

**CRISTIAN POMOHACI**

**DANIELA PÂRLEA**

**ANALIZA DATELOR**

Ediția a II-a

© Editura Fundației *România de Mâine*, 2008

**Editură acreditată** de Ministerul Educației, Cercetării și Tineretului  
prin Consiliul Național al Cercetării Științifice din Învățământul Superior

**Descrierea CIP a Bibliotecii Naționale a României**  
**POMOHACI, CRISTIAN MIHAI**

**Analiza datelor.** Ediția a II-a / Cristian M. Pomohaci,  
Daniela Pârlea. – București, Editura Fundației *România de*  
*Mâine*, 2008

Bibliogr.

ISBN 978-973-163-162-2

303.442.22(075.8)

Reproducerea integrală sau fragmentară, prin orice formă și prin orice  
mijloace tehnice, este strict interzisă și se pedepsește conform legii.

*Răspunderea pentru conținutul și originalitatea textului revine exclusiv*  
*autorului/autorilor*

**UNIVERSITATEA *SPIRU HARET***  
**FACULTATEA DE SOCIOLOGIE-PSIHOLOGIE**

**CRISTIAN POMOHACI**

**DANIELA PÂRLEA**

# **ANALIZA DATELOR**

**Ediția a II-a**

**EDITURA FUNDAȚIEI ROMÂNIA DE MÂINE**  
**București, 2008**



## CUPRINS

|                          |   |
|--------------------------|---|
| <i>Introducere</i> ..... | 9 |
|--------------------------|---|

### 1. NOȚIUNI PRELIMINARE

|                                                               |    |
|---------------------------------------------------------------|----|
| 1.1. Cunoașterea .....                                        | 11 |
| 1.2. Cum putem defini Analiza datelor? .....                  | 12 |
| 1.3. Stabilirea ipotezelor .....                              | 17 |
| 1.3.1. Caracteristici ale ipotezelor .....                    | 17 |
| 1.3.2. Erori frecvente în formularea ipotezelor .....         | 18 |
| 1.4. Grupul de subiecți .....                                 | 19 |
| 1.5. Ce metode putem folosi pentru prelucrarea datelor? ..... | 23 |
| 1.5.1. Metoda observației sociologice .....                   | 24 |
| 1.5.2. Ancheta bazată pe chestionar .....                     | 26 |
| 1.5.3. Testele psihologice .....                              | 31 |
| 1.5.4. Sociometria .....                                      | 32 |
| 1.6. Măsurarea .....                                          | 33 |

### 2. REPREZENTAREA DATELOR CA MOD DE EXAMINARE PRELIMINARĂ A FENOMENULUI STUDIAT

|                                                        |    |
|--------------------------------------------------------|----|
| 2.1. Serii de numere .....                             | 38 |
| 2.2. Gruparea datelor .....                            | 41 |
| 2.3. Reprezentări grafice ale seriilor de numere ..... | 42 |
| 2.3.1. Diagramele de tip linie .....                   | 43 |
| 2.3.2. Histogramele .....                              | 48 |
| 2.3.3. Alte tipuri de diagrame .....                   | 50 |
| 2.3.4. Sociograme .....                                | 56 |
| 2.3.5. Alte reprezentări ale datelor .....             | 58 |

|                                                                        |    |
|------------------------------------------------------------------------|----|
| 2.4. Noțiuni introductive de prezentare a datelor în Power Point ..... | 59 |
| 2.4.1. Ce este Power Point? .....                                      | 59 |
| 2.4.2. Cum începem? .....                                              | 60 |
| 2.4.3. Cum realizăm vizualizarea în Power Point? .....                 | 62 |
| 2.4.4. Modul de vizualizare Slide-Show .....                           | 63 |
| 2.5. Graficul Gantt .....                                              | 63 |
| 2.6. Concluzii .....                                                   | 64 |

### 3. STATISTICA ÎN CERCETAREA SOCIALĂ

|                                                                                                              |     |
|--------------------------------------------------------------------------------------------------------------|-----|
| 3.1. Caracteristici (variabile) .....                                                                        | 65  |
| 3.2 Populație .....                                                                                          | 67  |
| 3.3. Eșantion. Lot .....                                                                                     | 70  |
| 3.4. Eveniment .....                                                                                         | 71  |
| 3.5. Minimul și maximum .....                                                                                | 74  |
| 3.6. Media .....                                                                                             | 75  |
| 3.7. Cuantile .....                                                                                          | 78  |
| 3.7.1. Mediana .....                                                                                         | 78  |
| 3.7.2. Cuartilele .....                                                                                      | 82  |
| 3.7.3. Valoare modală .....                                                                                  | 83  |
| 3.8. Indicatori ai dispersiei .....                                                                          | 84  |
| 3.8.1. Amplitudinea .....                                                                                    | 85  |
| 3.8.2. Abaterea medie ( $A_m(a)$ ) .....                                                                     | 85  |
| 3.8.3. Varianța .....                                                                                        | 87  |
| 3.8.5. Coeficient de Omogenitate (variabilitate) .....                                                       | 88  |
| 3.9. Compararea mediilor .....                                                                               | 91  |
| 3.9.1. Cum comparăm media unui grup cu o valoare dată .....                                                  | 91  |
| 3.9.2. Testarea ipotezei privind diferența dintre mediile a două<br>grupuri cu număr redus de subiecți ..... | 92  |
| 3.10. Testul $\chi^2$ .....                                                                                  | 94  |
| 3.11. Coeficientul de corelație .....                                                                        | 97  |
| 3.12. Regresia liniară .....                                                                                 | 101 |
| 3.13. Rezolvare exerciții din capitolul 3 .....                                                              | 104 |

### 4. NOȚIUNI INTRODUCTIVE DESPRE GRAFURI ȘI BAZE DE DATE

|                                      |     |
|--------------------------------------|-----|
| 4.1. Grafuri .....                   | 116 |
| 4.1.1. Scurt istoric .....           | 116 |
| 4.1.2. Definiția unui graf .....     | 117 |
| 4.1.3. Clasificarea grafurilor ..... | 119 |

|                                                                                     |     |
|-------------------------------------------------------------------------------------|-----|
| 4.1.4. Arbori .....                                                                 | 120 |
| 4.1.5. Proprietăți ale grafurilor .....                                             | 122 |
| 4.1.6. Altă posibilitate de studiu al relațiilor dintr-un grup<br>de indivizi ..... | 124 |
| 4.1.7. Reprezentarea matriceală a unui graf .....                                   | 125 |
| 4.2. Baze de date .....                                                             | 127 |

## 5. UTILIZAREA EXCEL ȘI SPSS ÎN STATISTICĂ

|                                                             |     |
|-------------------------------------------------------------|-----|
| 5.1. Utilizarea Excel în statistică .....                   | 129 |
| 5.1.1. Inserarea unei funcții .....                         | 130 |
| 5.1.2. Minim și maxim dintr-un șir de date .....            | 132 |
| 5.1.3. Media .....                                          | 134 |
| 5.1.4. Mediana .....                                        | 135 |
| 5.1.5. Amplitudinea .....                                   | 136 |
| 5.1.6. Coeficientul de omogenitate .....                    | 137 |
| 5.1.7. Compararea mediei unui grup cu o valoare dată .....  | 138 |
| 5.1.8. Testul $t$ .....                                     | 140 |
| 5.1.9. Testul $\chi^2$ .....                                | 142 |
| 5.1.10. Coeficientul de corelație .....                     | 144 |
| 5.1.11. Regresia liniară .....                              | 146 |
| 5.1.12. Tabel funcții Excel pentru calcule statistice ..... | 149 |
| 5.2. Utilizarea SPSS .....                                  | 150 |
| 5.2.1. Ferestre și fișiere în SPSS .....                    | 151 |
| 5.2.2. Introducerea datelor .....                           | 151 |
| 5.2.3. Crearea unui fișier de date .....                    | 153 |
| 5.2.4. Statistica descriptivă a bazei de date .....         | 156 |
| 5.2.5. Corelația .....                                      | 158 |
| 5.2.6. Compararea între mediile a două loturi .....         | 161 |
| 5.2.7. Testul $\chi^2$ (hi-pătrat) .....                    | 167 |
| 5.2.8. Coeficientul alpha .....                             | 173 |
| 5.3. Concluzii .....                                        | 176 |



## INTRODUCERE

*Acest curs reprezintă o încercare sinteză a unor studii realizate, până în acest moment, în domeniul analizei datelor. În principal, am urmărit sistematizarea unor informații care să ajute studentul în înțelegerea și realizarea unui proiect de diplomă ca un prim pas în realizarea cercetării științifice.*

*În general, pentru studenții de la științele sociale este destul de complicată însușirea unor noțiuni ce au un caracter mai tehnic. Aceasta se datorează anumitor abordări defectuoase a materiilor tehnice. Nu o dată s-a întâmplat la cursul de Analiza datelor ca o singură formulă pusă pe tablă să creeze o reacție negativă din partea studenților. De aceea, vom încerca o abordare mai „umană” a acestei materii, pornind de la experiențele pe care le-am avut în decursul anilor de predare. Lucrarea este structurată pe 5 capitole. Succesiunea capitolelor a fost impusă de dorința de a grada prezentarea de la ușor la noțiuni mai dificile, pentru a face mai accesibilă parcurgerea cărții.*

*În primul capitol, se dă definiția disciplinei de Analiza Datelor, rolul și locul ei în cercetarea sociologică și psihologică. Apoi se prezintă o parte din etapele realizării unui proiect de cercetare.*

*A doua parte reia anumite noțiuni din cartea Informatică utilizată în sociologie și psihologie, noțiuni legate de prezentarea datelor: realizarea de diagrame, de scheme și despre Microsoft Power Point.*

*În capitolul trei se prezintă intuitiv câteva din instrumentele statistice utilizabile în cercetare.*

*În capitolul 4 se prezintă instrumente de stocare a rezultatelor, adică noțiuni despre baze de date și grafuri, noțiuni ce vor fi utile prezentării de SPSS-ului și a Excel-ului.*

*Odată reamintite noțiunile de statistică se va trece la capitolul 5 în care se vor prezenta noțiuni de realizare a calculelor statistice cu ajutorul SPSS-ului și a Excel-ului.*

*Se consideră cunoscute, și deci nu se vor mai relua, noțiuni de utilizarea computerului, noțiuni de metode și tehnici de cercetare socială (deși o parte din acestea se vor relua pe scurt).*

# 1. NOȚIUNI PRELIMINARE

## 1.1. Cunoașterea

Ce ar putea determina pe cineva să citească o carte? Să deschidă cartea și să vadă ce poate fi scris în paginile ei? „Curiozitatea”, au răspuns o mare parte din studenți când le-am pus această întrebare. Curiozitatea este o manifestare a unei întrebări pe care ne-o punem: putem găsi ceva aici? Putem cunoaște ceva? Deci curiozitatea ar putea fi o manifestare a dorinței de cunoaștere. Cunoașterea este motorul care ne duce în fiecare zi dintr-o parte în alta. Vrem să știm cum va fi vremea, deschidem radioul și așteptăm să vină știrile despre vreme. Vedem un meci la televizor, apoi a doua zi cumpărăm ziarul ca să vedem, să cunoaștem și cum au văzut alții, sau poate ce au văzut alții și noi n-am văzut la acel meci. Deci, faptul că ceva s-a întâmplat nu e suficient, dorința noastră este de a analiza, de a obține cât mai multe date despre ceea ce ne interesează. Totuși, din multitudinea de lucruri care se întâmplă în jurul nostru cum alegem pe unele și respingem pe altele? Cum alegem să cunoaștem unele și să nu cunoaștem altele? În cartea ei<sup>1</sup>, Ruane determină mai multe căi de cunoaștere:

- Cunoașterea tradițională
- Cunoașterea dată de autorități
- Cunoașterea dată de bunul-simț
- Cunoașterea dată de intuiție
- Cunoașterea realizată prin metode științifice

Faptul că dintre cele cinci tipuri de cunoaștere cea mai îndepărtată de subiectivism este cunoașterea realizată prin metode științifice, dă o motivație acestei cărți. Specificăm totuși că nu trebuie

---

<sup>1</sup> Vezi la bibliografie [Rua].

respinse celelalte tipuri de cunoaștere, dar acestea nu constituie subiectul manualului de față.

## 1.2. Cum putem defini Analiza Datelor?

În cele ce urmează vom prezenta două posibile definiții ale analizei datelor.

Într-o primă prezentare ne bazăm pe definirea succesivă a unor concepte, construcția finalizându-se cu definirea analizei datelor. La baza analizei datelor stă conceptul de *mesaj*. *Mesajul* reprezintă baza comunicării, voluntare sau involuntare, între un emițător și un receptor. Ceea ce acceptă un receptor se numesc *date*. Atragem atenția asupra faptului că nu ceea ce dorește să transmită emițătorul poartă numele de *date*, deoarece o parte din ceea ce se transmite se poate pierde pe drumul între emițător și receptor. Conștientizarea prin analizare și prelucrarea datelor receptate transformă datele în *informație*. Deci, o posibilă definiție a analizei datelor ar fi: *Procesul prin care datele primite de receptor se transformă în informație*.

Schematic aceasta se poate reprezenta astfel:

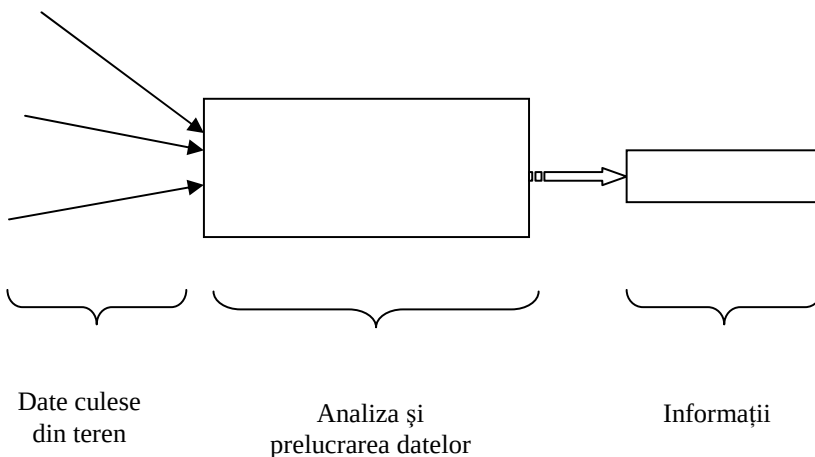


Fig. 1. Trecerea de la date la informații cu ajutorul Analizei Datelor

Informațiile odată dobândite devin parte a cunoștințelor. De aceea, putem spune că procesul de finalizare a analizei și de prelucrare a datelor constă în transformarea datelor în informații, iar informațiile au un rol major în formarea de cunoștințe. Ce se înțelege prin *cunoștințe*? „Cunoștințele sunt elemente abstracte și individuale despre obiectele din lumea reală, însușite prin educație și experiență.”<sup>2</sup>

Am insistat pe o prezentare în detaliu a acestor noțiuni pentru că ele reiau, pe scurt, unele din principiile oricărei cercetări.

### **Exemplul 1**

Acum câțiva ani, un student m-a rugat să-l ajut să facă o repetiție pentru susținerea proiectului de diplomă. Era vorba de un studiu comparativ privitor la relația dintre mamă și copil între mamele care aveau program de 8 ore la serviciu și cele care aveau program de 4 ore. A început să vorbească despre tema pe care și-o alesese la proiect. A prezentat obiectivele, ipotezele, instrumentele de lucru, chiar și niște diagrame destul de sugestive. În final, a văzut că între cele două loturi erau diferențe semnificative. „Ce părere aveți?” m-a întrebat. „Niciuna, pentru că încă nu ai terminat”, i-am spus. „Ceea ce mi-ai prezentat până aici sunt doar niște date cu o prelucrare preliminară.”

Prin prisma conceptelor prezentate în exemplul de mai sus, vom spune că studentul este un receptor al unor date culese din teren; prin prelucrarea primară a datelor, el a transformat datele, în informație, dar informația, deși corectă, între cele două loturi fiind diferențe semnificative, totuși era insuficientă pentru a fi parte integrantă a unor cunoștințe despre fenomenul studiat.

Cea de-a doua metodă pornește de la definiția din dicționar a celor două cuvinte care compun sintagma *Analiza datelor*. Conform

---

<sup>2</sup> Cf. [Vel] .

*Dicționarului Explicativ al Limbii Române\** pentru a analiza găsim următoarea definiție:

„ANALIZĂ, analizez, vb. I. Tranz. 1. A cerceta un întreg, un fenomen etc., examinând fiecare element în parte. ♦ A examina un text din diferite puncte de vedere.” sau „ANALIZĂ vb. 1. a cerceta, a examina, a investiga, a studia, a urmări, (livr.) a considera, (înv.) a medita, a privi, a socoti, (fig.) a explora, (înv. fig.) a scărmana. (~ cauzele unui fenomen.) 2. v. examina. 3. a comenta, a explica, a interpreta, a tâlcului, (înv.) a întoarce, (fig.) a descifra. (~ un text literar.)”

Să luăm, pe rând, elementele ce apar în această definiție: o cercetare pornește de la un fenomen, de la un **întreg** care ne atrage atenția și prima reacție este să privim fiecare detaliu al lucrului analizat, adică **să examinăm fiecare element în parte**. Cercetarea poate porni de la un text la care, mai întâi, ne informăm asupra diferitelor puncte de vedere asupra textului, urmând să încercăm un punct de vedere diferit.

Pentru *date* găsim următoarea definiție:

„DAT, -Ă, dați, -te, adj., s.f. I. Adj. Pus la dispoziție, oferit; înmănat, transmis, prezentat, dăruit. ♦ Expr. La un moment dat = într-un anumit moment; în clipa aceea. În cazul dat = în acest caz, în cazul de față. Dat fiind (că...) = ținând seama de... (sau că...), având în vedere (că...); deoarece. [...] Dat uitării = uitat, părăsit. II. S.f. 1. Momentul, împrejurarea (repetabilă) când se produce un fapt; oară, rând. ♦ Loc. adv. Data trecută = cu prilejul anterior. Data viitoare = într-o împrejurare ulterioară. De data aceasta (sau asta) ori de astă dată = de rândul acesta, acum. Pe dată ce... (sau cum...) = îndată; pe loc, numaidecât. O dată = într-un singur caz. Nu o dată = de multe ori. Încă o dată = din nou. 2. (La pl.) Fapte stabilite (de știință), elemente care constituie punctul de plecare în cercetarea unei probleme, în luarea unei hotărâri etc. III. S.f. (Reg.) Soartă, destin. ♦ Cum (sau precum) e data = după cum e obiceiul, datina. – V. da2.”

---

\* Academia Română, Institutul de Lingvistică „Iorgu Iordan”, *Dicționarului Explicativ al Limbii Române*, Ediția a II-a, Editura Univers Enciclopedic, București, 1998.

Pornind de la aceste două definiții putem spune că **Analiza Datelor** este disciplina care se ocupă cu *cercetarea, examinarea, investigarea, interpretarea faptelor stabilite științific, fapte care constituie punctul de plecare în cercetarea unei probleme, în luarea unei hotărâri*.

Deci, cu alte cuvinte, putem spune că analiza datelor reprezintă o etapă în cercetarea științifică a unui fenomen.

În continuare vom discuta despre etapele cercetării. Acestea sunt:

1. Proiectarea
2. Fixarea obiectivelor, stabilirea ipotezelor
3. Selecția grupului/grupurilor de subiecți
4. Alegerea metodelor ce urmează să fie folosite în culegerea și prelucrarea datelor (atât a instrumentelor specifice, cât și a celor statistice)
5. Alegerea lotului de subiecți (în cazul sondajelor de opinie este vorba de designul eșantionului reprezentativ)
6. Analiza Datelor
7. Diseminarea rezultatelor (prezentarea proiectului de licență în fața comisiei de examinare – în cazul absolvenților, prezentarea rezultatelor finanțatorului, publicarea rezultatelor în reviste de specialitate etc.)

Toate aceste etape ale cercetării sunt importante deoarece orice eroare în gestionarea lor poate duce la erori destul de mari în analiza datelor.

Proiectul de diplomă reprezintă o primă ocazie pentru student de a fi pus în situația unei cercetări pe care trebuie s-o realizeze și s-o prezinte singur în fața unei comisii de examinare. De aceea, considerăm că este util, în vederea pregătirii proiectului, să prezentăm câteva noțiuni de management de proiect.

De cele mai multe ori studenții pornesc la realizarea cercetării, ce va sta la baza proiectului de diplomă, „construiesc” această cercetare fără a-și planifica timpul sau resursele, ceea ce are ca efect faptul că ajung în criză de timp.

Ce înseamnă a realiza un proiect? „Un proiect reprezintă [...] un grup de activități relaționate, în mod organizat, pentru îndeplinirea unui scop”<sup>3</sup>.

În realizarea oricărui proiect trebuie să avem în vedere trei constrângeri principale<sup>4</sup>:

- timp (durata efectivă de realizare a cercetării)
- resurse (materiale, financiare etc.)
- specificitate a rezultatelor.

De aceea, ar fi bine ca studenții să țină cont de aceste trei constrângeri când își planifică activitățile legate de realizarea unui proiect. Planificarea activităților se realizează în etape sau subetape pentru procesele mai complicate. Atunci când proiectăm aceste etape există două moduri de abordare:

1. *planificarea directă* – de la prima etapă la etapa finală (aceasta însemnând o cunoaștere foarte bună a tehnicilor cercetării)
2. *planificarea inversă*<sup>5</sup> – pornim dinspre etapa finală spre prima etapă (în acest caz este necesar să vizualizăm foarte bine etapele pentru a găsi drumul optim dintre ultima și prima etapă).<sup>6</sup>

Un nou mod de abordare a proiectării unei cercetări este cel realizat cu ajutorul tehnicii de programare în rețea. Din aceste tehnici amintim:

- graficul Gantt
- graficul PERT<sup>7</sup>
- graficul CPM<sup>8</sup>

---

<sup>3</sup> Cf. [McC] p. 29.

<sup>4</sup> Cf. [McC].

<sup>5</sup> *Backward planning* – denumire preluată din lucrarea [McC].

<sup>6</sup> Pentru studentul ce-și pregătește proiectul de diplomă recomandăm cea de-al doilea tip de proiectare, *backward*, întrucât există constrângerea temporală.

<sup>7</sup> **P**rogram **E**valuation and **R**evue **T**echnique.

<sup>8</sup> Metoda Drumului Critic.

La finalul capitolului 2 vom prezenta câteva noțiuni legate de graficul Gantt.

În finalul acestei secțiuni revenim la problema gestionării timpului. Kerzner dă o listă a activităților ce pot „răpi timp” în realizarea unui proiect<sup>9</sup>. Din acestea amintim:

- munca incompletă
- amânarea nejustificată în luarea unor decizii
- folosirea exagerată a telefonului, chat-ului etc.
- întârzierile la întâlniri
- corectarea superficială a unor erori
- planificarea pe termen scurt (lipsa unei planificări pe termen lung)
- perfecționismul
- schimbări dese ale planificării
- analizele pe prea multe niveluri

### 1.3. Stabilirea ipotezelor

Odată aleasă tema și stabilite obiectivele, problema ce trebuie rezolvată este cea a stabilirii ipotezelor. Atunci când se stabilesc ipotezele, trebuie avute în vedere două repere de bază: fiecare ipoteză să provină dintr-un obiectiv, și, în același timp, să fie în acord cu instrumentele utilizate. Un alt aspect definitoriu al ipotezelor îl reprezintă faptul că ipotezele presupun „o solidă informație în domeniu, care oferă și *cadrul conceptual* al ipotezei”<sup>10</sup>.

#### 1.3.1. Caracteristici ale ipotezelor

I. Radu prezintă anumite caracteristici ale ipotezelor. Din acestea amintim<sup>11</sup>:

- Explică un ansamblu de date
- Ipoteza trebuie să fie „bine întemeiată, *plauzibilă* sau *verosimilă*”
- Ipotezele trebuie să fie verificabile

---

<sup>9</sup> Cf. [McC] p. 199.

<sup>10</sup> [Rad] p. 21.

<sup>11</sup> [Rad] p. 22.

### 1.3.2. Erori frecvente în formularea ipotezelor

În cele ce urmează vom prezenta câteva din erorile întâlnite în formularea unor ipoteze:

1. Exprimarea mult prea generală a ipotezei

#### **Exemplul 2**

*Hiperemotivitatea conduce la anxietate ridicată și depresie.*

2. folosirea gradelor de comparație, fără a spune nimic (sau a folosi o exprimare neclară), privitoare la termenii ce se vor compara

#### **Exemplul 3**

*Există o alegere din partea angajaților, începând de la un anumit nivel al salariului, între sporirea numărului de ore lucrate și sporirea timpului liber*

3. formularea prin negare

#### **Exemplul 4**

*Succesul școlar nu are drept determinant principal inteligența peste medie*

4. folosirea de adjective necuantificabile

#### **Exemplul 5**

*Copiii expuși, în mai mare măsură, la pericolul de a fi supuși unor **rele** tratamente se regăsesc, în special, în familiile cu situație materială precară*

5. folosirea de cuvinte-balast

#### **Exemplul 6**

***Indiferent** de tipul de abuz suferit, copiii-victime prezintă **totuși** un tablou, simptomatologic comun: relaționarea defectuoasă, scăderea randamentului școlar (**eventual abandon școlar**), stare agresivă și comportamente agresive*

## 6. exprimarea de concluzii

### **Exemplul 7**

*Cu cât Biserica va blama homosexualitatea, cu atât atitudinea românilor va fi mai negativă (**România este o țară unde sentimentul religios este foarte puternic, iar Cuvântul Bisericii are un rol decisiv**)*

Desigur, acestea sunt doar o parte din erorile frecvente în formularea ipotezelor, dar am ținut să atragem atenția asupra lor deoarece formularea unei ipoteze reprezintă unul din pașii cei mai importanți ai cercetării.

O altă problemă, legată de ipoteze, o reprezintă „**clișeul confirmării ipotezei de start**”<sup>12</sup>. Aceasta se referă la faptul că, de multe ori, mai ales la cei care sunt la începutul activității de cercetare, există tendința ca lucrarea să se axeze pe confirmarea cu orice preț a ipotezei de start, ignorându-se discordanțele cu aceasta.

### **Exemplul 8**

S-a aplicat o sociogramă într-o clasă. În urma aplicării sociogramei s-a determinat liderul. Analizând însă rezultatele, s-a observat faptul că doi elevi din clasă, deși ei și-au exprimat opiniile, nu au primit niciun punct de la colegi. Ceea ce ar putea duce la anumite întrebări privind integrarea lor în grup. Deci, pe lângă ipoteza de start, se poate adăuga, în această situație, și un studiu de caz pentru cei doi elevi.

## 1.4. Grupul de subiecți

Am folosit sintagma *grupul de subiecți* deoarece cele ce urmează sunt valabile atât pentru eșantion, cât și pentru lot.

În principal, în această secțiune ne vom ocupa de experimentele comparative. Pentru a realiza un astfel de studiu, trebuie să alegem un grup de control și un grup experimental. Grupul

---

<sup>12</sup> [Rad], p. 225.

de control este acel grup care nu este supus stimulilor specifici experimentului, este un grup care propriu-zis nu intervine în experiment. De unde și denumirea de *grup de martor*.

### **Exemplul 9**

S-a realizat o cercetare în care s-a pus problema dacă relaxarea, ca tehnică psihoterapeutică, conduce la ameliorarea simptomelor nevrotice. Pentru aceasta s-a luat ca grup experimental 30 de pacienți, din care 15 bărbați și 15 femei. Aceștia au fost selecționați, în mod aleator, dintre persoanele cu tulburări anxioase ce s-au prezentat la clinică, solicitând psihoterapie. Grupul martor conține 30 de persoane selectate în mod aleator, dintre care 15 bărbați și 15 femei care prezentau același tip de tulburări anxioase ca și cei din grupul experimental. Se poate observa că puteam să lucrăm cu un singur grup de subiecți unde, înainte de a i se aplica tehnicile psihoterapeutice, era grupul martor și după aplicarea tehnicilor terapeutice devine grup experimental.

Atunci când efectuăm măsurători asupra unui grup trebuie să ținem seama de o serie de factori<sup>13</sup> cum ar fi:

- **Efectul de maturare** – diferențele ce ar putea apărea între două măsurători să nu se datoreze stimulilor aplicați în cadrul experimentului, ci maturării ce a avut loc pe parcursul experimentului

### **Exemplul 10**

La clasa a treia se dorește să se studieze dacă prin folosirea unor tehnici de dezvoltare a atenției elevii se descurcă mai bine la efectuarea calculelor de înmulțire. După un semestru se compară rezultatele de la începutul anului cu cele de la sfârșitul

---

<sup>13</sup> Cf. [Rad] .

semestrului. Se observă că majoritatea elevilor au obținut scoruri mai bune. În acest caz, însă nu putem folosi ca grup martor grupul de la început și ca grup experimental grupul la sfârșitul semestrului, deoarece nu putem ști dacă evoluția pozitivă se datorează metodei aplicate sau faptului că elevii în timp, lucrând la matematică, oricum și-ar fi îmbunătățit performanțele. În acest caz, se recomandă folosirea a două grupuri distincte: un grup de elevi cărora li se aplică tehnicile de dezvoltare a atenției – grup experimental și un grup de elevi cărora nu li s-a aplicat nicio tehnică de dezvoltare a atenției – grup martor. Comparatia se va face între aceste grupuri atât la începutul semestrului, cât și la sfârșitul semestrului. Grupurile trebuie alese la începutul semestrului astfel încât să nu fie diferențe semnificative între ele, iar la sfârșit eventuala apariție a unei diferențe ne poate aduce informații despre efectele tehnicilor de dezvoltare a memoriei asupra copiilor.

- ***Efectul testării repetate*** – aplicând la intervale destul de scurte același test, de mai multe ori, pe același lot de subiecți rezultatele vor fi influențate și de faptul că subiecții ajung să cunoască foarte bine testul

### ***Exemplul 11***

Măsurăm anxietatea zilnic, timp de o săptămână înainte de operație, la pacienții din același salon. Diferențele care ar putea apărea în acest caz se pot datora și faptului că pacienții se obișnuiesc cu testul și dau alte răspunsuri pentru a nu se repeta.

- ***Influența evenimentelor externe*** – apariția unui eveniment extern experimentului, eveniment ce poate influența subiecții.

### **Exemplul 12**

Se aplică un chestionar pe problema corupției în mediul urban și în mediul rural. În prima zi se aplică în mediul rural. Seara, la televizor, la știri se vorbește despre un proces în care există acuzația de corupție a unui demnitar. Diferențele ce ar putea apărea, a doua zi, la aplicarea chestionarului în mediul urban se pot datora și influenței emisiunii din seara de dinaintea aplicării chestionarului

- **Corectitudinea selecției subiecților** – selecția subiecților trebuie să se realizeze pe cât posibil independent de voința subiecților și trebuie evitat subiectivismul celui care face selecția

### **Exemplul 13**

Într-o școală s-a aplicat un test de anxietate doar la elevii ce s-au oferit voluntari în a completa testul. S-au obținut valori scăzute ale anxietății. Totuși, nu putem realiza o interpretare a datelor deoarece nu știm dacă aceste valori scăzute, ce reprezintă realitatea sau faptul că lotul a fost compus din elevi ce s-au oferit voluntari, au dus la valori scăzute ale anxietății.

### **Exemplul 14**

S-a aplicat un chestionar în care se studia interesul populației dintr-un oraș față de problemele din învățământ. Chestionarul a fost aplicat în curtea unei școli printre părinții ce așteptau copiii de la clasele primare să iasă de la cursuri. Rezultatele acestui chestionar nu pot fi validate, deoarece ceea ce s-a obținut în urma prelucrării nu se poate spune dacă este răspunsul populației față de întrebările puse sau e vorba de influența faptului că s-a aplicat printre părinții elevilor din ciclul primar.

- **Riscul de „a pierde” subiecți**<sup>14</sup> – din cauza unor condiții ce nu depind de subiect – o mare parte din subiecți nu răspund la toate întrebările.

### **Exemplul 15**

Se dorește să se vadă impactul unui film asupra spectatorilor. Se aplică la o grădină unde rulează acel film. Se aplică un chestionar înainte de a rula filmul, urmând, ca la final, să se mai aplice chestionarul încă o dată. Pe la mijlocul filmului însă începe ploaia și toți spectatorii au plecat în grabă astfel încât operatorul nu a mai apucat să aplice și a doua oară chestionarul.

Desigur, există mai multe aspecte care trebuie urmărite atunci când se selectează un grup în vederea realizării unui experiment. Am prezentat doar situațiile cele mai des întâlnite. Pentru a vedea și alte cazuri recomandăm, de exemplu, consultarea [Rad].

### **1.5. Ce metode putem folosi pentru prelucrarea datelor?**

În această secțiune dorim să realizăm o trecere în revistă a principalelor instrumente ce se folosesc în prelucrarea datelor, altele decât cele statistice<sup>15</sup>. Chiar cu riscul de a fi considerați negativiști la metodele clasice de cercetare vom începe cu prezentarea neajunsurilor și apoi a avantajelor. Alegem aceasta deoarece, de cele mai multe ori, în utilizarea metodelor, studenții se concentrează pe avantajele acelui instrument, ignorând dezavantajele ce le pot apărea chiar și în condițiile folosirii corecte a instrumentului.

---

<sup>14</sup> În [Rad] acest caz poartă denumirea de „moarte experimentală”, p. 39.

<sup>15</sup> De instrumentele statistice ne vom ocupa în capitolul 3.

### 1.5.1. Metoda observației sociologice

Una din metodele de bază în sociologie o constituie *observația*. Având în vedere că această metodă se află la granița dintre intuiție și raționament, folosirea ei trebuie realizată cu foarte mare grijă. În primul rând, fără o susținere din partea altor metode și instrumente, observația poate deveni sursă de eroare în analiza datelor. Dintre alte neajunsuri ale acestei metode amintim:

- necesită timp îndelungat de folosire
- se folosește pe eșantioane<sup>16</sup> mici
- la nivel de observație nu se recomandă încercarea de a identifica relații cauzale

Și atunci apare întrebarea: „De ce să se folosească această metodă?” În primul rând, această metodă este utilă atunci când e folosită împreună cu alte metode. Printre avantajele acestei metode avem:

- permite accesul direct și nemijlocit al cercetătorului la realitatea socială
- asigură un grad mare de obiectivitate și complexitate a imaginii rezultate
- ne ajută să înregistrăm fenomenul studiat în condițiile sale proprii de desfășurare
- poate realiza atât înregistrarea comportamentelor individuale, cât și a celor colective

Putem distinge două mari tipuri de observație:

Un prim mare tip de observație îl constituie *observația spontană*. Aceasta reprezintă genul de observație, realizată la nivelul simplului contact al cercetătorului cu realitatea. Dintre dezavantajele acestui gen de observație am aminti:

- se realizează, în general, pe o bază subiectivă
- rareori se poate repeta în aceleași condiții
- înregistrarea se face pe baza memoriei, deci există pericolul trunchierii informațiilor

---

<sup>16</sup> Despre noțiunea de *eșantion* vezi capitolul 3, secțiunea 3.3.

*Avantaje:* acest gen de observație este foarte util în găsirea de direcții noi de cercetare, în căutarea de noi soluții la probleme de rezolvat. Unul din exemplele clasice este descoperirea penicilinei de către Flemming. Totuși, continuând analogia, fără să trecem la *observația științifică*, observația spontană, care i-a atras atenția lui Flemming, ar fi avut rezultate neconcludente<sup>17</sup>.

Despre *observația științifică*, Henri H. Stahl<sup>18</sup> atrăgea atenția asupra următoarelor note definitorii:

- „[...] are la bază o concepție științifică despre lumea înconjurătoare
- [...] vizează totalitatea aspectelor fenomenului studiat
- [...] este condusă după anumite reguli
- [...] asigură un grad mare de obiectivitate în aprecierea fenomenelor
- [...] procedează la descompunerea obiectului observat în elementele lui componente și apoi trece la evidențierea întregului
- datele și informațiile rezultate în urma observației sunt înregistrate într-o fișă de observație, în vederea prelungirii lor ulterioare”

Totuși trebuie remarcat că observația științifică devine mult mai bogată în rezultate atunci când reprezintă o urmare a unei observații spontane.

Desigur, acesta nu este singurul mod de a clasifica observațiile. Am putea aminti că observațiile se pot clasifica după *gradul de structurare*, după *gradul de implicare al cercetătorului* în sistemul studiat, *în funcție de durata observației*, *natura observației*, *gradul de extensiune și de profunzime al observației*<sup>19</sup>. Acestea sunt tratate, pe larg, în manualele de specialitate din care putem să recomandăm [Cau].

---

<sup>17</sup> Cel mult s-ar fi gândit ca altă dată să fie mai atent și să nu mai lase laboratorul cu ustensilele murdare.

<sup>18</sup> Apud [Cau], p. 103-104.

<sup>19</sup> Cf. [Cau].

Pentru a fi siguri că folosim corect metoda observației sociologice trebuie să ne facem niște reguli pe care să le respectăm. Desigur, nu există reguli general valabile, dar, dacă s-ar încerca găsirea unor elemente comune, un punct de plecare ar fi aprecierile lui T. Calpow<sup>20</sup>:

- observația trebuie să aibă la bază o foarte bună pregătire teoretică a observatorului
- formularea, cu precizie, a tehnicilor de observare și a procedeele de notare a observațiilor
- notarea faptelor de observație, pe cât posibil la fața locului, pe teren
- trebuie realizată distincția dintre faptele observate și părerile observatorului în sensul că acestea trebuie notate separat
- observarea trebuie să se realizeze continuu și sistematic, avându-se în vedere urmărirea obiectivelor precis formulate.

Odată luate aceste precauții, prin folosirea metodei observației sociologice, putem ajunge la cunoașterea și înțelegerea realității, cât și la explicarea celor observate.

#### 1.5.2. *Ancheta bazată pe chestionar*

Atunci, când dorim să explicăm comportamentele umane și să identificăm factorii ce determină aceste comportamente, un instrument foarte util îl reprezintă *chestionarul*. Când se aplică un chestionar există mai mulți factori care pot influența răspunsurile. Dintre aceștia amintim: personalitatea celui care anchetează, personalitatea celui anchetat, tema anchetei, locul unde se aplică chestionarul, timpul de desfășurare al chestionarului.

Chestionarul reprezintă un instrument al unui tip de anchetă. De aceea, acest tip de anchetă mai poartă și denumirea de anchetă prin chestionar. Dintre caracteristicile anchetei prin chestionar am aminti:

---

<sup>20</sup> Apud [Che1].

1. *Caracterul standardizat* – numărul întrebărilor, ordinea întrebărilor și forma lor este stabilită clar dinainte. Tot cadrul standardizat al anchetei este asigurat de faptul că se stabilește dinainte numărul persoanelor alocate fiecărei caracteristici<sup>21</sup>, acest număr rămânând neschimbat pe tot parcursul anchetei.

2. *Tehnici specifice instrumentului folosit (chestionarului)*

3. *Numărul mare de persoane pe care se aplică chestionarul* – ancheta de chestionar, pentru a fi corect realizată, trebuie aplicată pe eșantioane<sup>22</sup> mari, existând o legătură strânsă între numărul de persoane pe care se aplică chestionarul și nivelul de semnificabilitate al rezultatelor.

Chestionarele se pot clasifica în funcție de trei criterii<sup>23</sup>:

- conținutul informațiilor obținute
  - chestionarele de date factuale
  - chestionarele de opinie
  - chestionarele speciale
  - chestionarele tip omnibus
- forma întrebărilor
  - chestionare cu întrebări închise
  - chestionare cu întrebări deschise
  - chestionare cu întrebări mixte
  - chestionare cu întrebări scalate
- modul de aplicare al chestionarelor
  - chestionare autoadministrate
  - chestionare administrate de către operatorii de anchetă

### **Exemplu 16**

Atunci când a avut de aplicat un chestionar, studentul a realizat un presondaj unde, printre alte întrebări, a pus și întrebarea deschisă: „Cum vi se par condițiile de la locul dumneavoastră de

---

<sup>21</sup> Sex, vârstă, mediu, nivel de studii etc.

<sup>22</sup> Despre noțiunea de *eșantion* vezi capitolul 3, secțiunea 3.3.

<sup>23</sup> Cf. [Che2].

*muncă?”* După centralizarea rezultatelor, deși sperase ca la această întrebare să poată determina anumite opțiuni pentru a face o interpretare, paleta de răspunsuri era foarte largă. Atunci, deoarece și-a dat seama că există riscul ca în cazul în care ar lăsa întrebarea deschisă interpretarea să fie mai greu de realizat, iar răspunsurile primite destul de greu de cuantificat, studentul a optat pentru transformarea acestei întrebări într-o întrebare închisă de forma:

*„Cât de mulțumit sunteți de condițiile de la locul dumneavoastră de muncă?”*

*a. mulțumit*

*b. nici mulțumit, nici nemulțumit*

*c. nemulțumit*

În general, chestionarul se administrează prin intermediul operatorilor pe teren. Desigur, mai există și alte moduri de aplicare a unui chestionar, cum ar fi: telefonic, prin scrisori, sau cu ajutorul Internetului (prin email, chat sau chiar pagini specializate în aplicarea electronică a răspunsului), dar acestea sunt mai puțin utilizate în comparație cu metoda clasică prin intermediul operatorilor de teren.

Elementul de bază al unui chestionar îl constituie întrebarea. Când analizăm o întrebare există cel puțin trei elemente importante la care facem referire:

- conținutul întrebării,
  - întrebări factuale
  - întrebări de cunoștințe
  - întrebări de motivație
- forma întrebării<sup>24</sup>,
- poziția întrebării în cadrul chestionarului
  - întrebări introductive
  - întrebări de conținut
  - întrebări de trecere
  - întrebările filtru

---

<sup>24</sup> Vezi clasificarea chestionarelor după forma întrebărilor.

- o întrebările bifurcate
- o întrebările de control
- o întrebări de identificare

În cele expuse până aici ne-am referit la chestionare doar din punctul de vedere al componentelor acestora. Dar ceea ce face un instrument să funcționeze este raportul ce se stabilește între diferitele părți ale instrumentului. Aceste raporturi conduc la evidențierea tehnicilor conform cărora se pot alcătui chestionarele<sup>25</sup>.

Dintre tehnicile de alcătuire a chestionarelor amintim: *tehnica pâlniei* – trecerea de la general la particular și *tehnica pâlniei răsturnate* – trecerea se face de la particular la general.

**Exemplul 17** (tehnica pâlniei)

1. Cum credeți că se trăiește în România comparativ cu acum un an?
2. Cum credeți că trăiți dumneavoastră comparativ cu acum un an?

**Exemplul 18** (tehnica pâlniei răsturnate)

1. În situații critice vă folosiți de autoritatea dumneavoastră de manager pentru a soluționa o problemă?
2. Când ar trebui un manager să se folosească de autoritatea dată de poziția lui în cadrul firmei?

Atunci când stabilim întrebările și structura chestionarului trebuie să avem în vedere posibile pericole ce pot apărea în gestionarea acestui instrument. Din acestea am aminti:

I. Atunci când într-un chestionar, datorită așezării întrebărilor, avem contaminarea răspunsurilor spunem că se manifestă *efectul de halo*.

---

<sup>25</sup> Cf. [Che2].

### Exemplul 19

1. Credeți că majoritatea politicienilor sunt corupți?
  - a. Da
  - b. Nu
  - c. Nu știu
2. Considerați că în sistemul sanitar există corupție?
  - d. Da
  - e. Nu
  - f. Nu știu
3. Se poate vorbi de corupție în sistemul juridic?
  - g. Da
  - h. Nu
  - i. Nu știu
4. Cât de mare considerați că este corupția în România?
  - j. Foarte mare
  - k. Mare
  - l. Medie
  - m. Mică
  - n. Foarte mică

II. Un alt posibil factor de eroare în utilizarea unui chestionar este *lungimea chestionarului*. S. Chelcea, de exemplu, recomandă un număr de 25 până la 50 de itemi<sup>26</sup>. Numărul de întrebări ce pot fi puse într-un chestionar reprezintă totuși o opțiune a celui care alcătuiește chestionarul.

III. O eroare destul de mare o constituie *erorile de eșantionare* care se pot manifesta sub mai multe forme. Dintre acestea amintim:

- Nerespectarea unor caracteristici ale populației de bază
- Numărul prea mare de refuzuri

---

<sup>26</sup> În [Che2].

IV. Un alt tip de erori îl reprezintă erorile datorate operatorilor de teren. Acestea pot fi:

- Personalitatea necorespunzătoare a operatorilor (datorată unei selecții incorecte a operatorilor) – aroganță, aspect fizic neplăcut, neseriozitate etc.
- Aplicarea incorectă a chestionarului – scurtarea întrebărilor, simplificarea excesivă a întrebării, schimbarea ordinii întrebărilor etc.
- Influențarea respondenților de către operatori pentru a da anumite răspunsuri
- Nerespectarea eșantionării sau a pasului de aplicare a chestionarului

V. O sursă de erori, deloc neglijabilă, o constituie cei care dau răspunsurile. Dintre acestea am aminti:

- Nesinceritatea răspunsurilor
- Refuzul de a completa sau de a participa
- Sensibilitatea subiecților la anumite teme
- Tendința de a da răspunsuri conforme cu ceea ce este de dorit din punct de vedere social
- Tendința de a se pune într-o lumină favorabilă

Deși aceste erori, expuse mai sus, sunt destul de frecvente, ancheta pe bază de chestionar rămâne un instrument important în cercetările sociale și, utilizat cu atenție, pentru a evita erorile expuse mai sus, poate constitui o importantă sursă de date.

### 1.5.3. *Testele psihologice*

Deoarece acest manual se adresează studenților ce sunt abia la începutul activității de cercetare, ne vom referi la teste, doar din punctul de vedere al utilizatorilor testelor<sup>27</sup>.

---

<sup>27</sup> Deși este laudabilă dorința unor studenți de a explora și a găsi noi metode și teste pe care mai apoi să le valideze, le recomandăm ca mai întâi să învețe să folosească foarte bine testele deja validate și după acumularea unei experiențe semnificative să poarte drumurile noi și neexplorate ale testării psihologice.

O primă problemă ce trebuie rezolvată atunci când se dorește folosirea unui test într-o cercetare este de a găsi testul potrivit subiectului studiat. Alegerea testului nu trebuie să se bazeze doar pe denumirea testului. Aceasta deoarece, așa cum remarca I. Radu<sup>28</sup>: „ (...) numele (...) spune puțin (...), el caută să ascundă obiectivul sau intenția testului pentru a evita răspunsul conformist ori favorabil (din partea subiectului)”. De aceea, recomandăm ca înaintea aplicării unui test să se parcurgă, cu atenție, manualul testului sau materialul de îndrumare ce însoțește testul.

Din cele expuse mai sus mai apare și un alt aspect al aplicării unui test. Cu cât subiectul căruia îi aplicăm testul știe mai puține despre test cu atât rezultatele vor fi mai aproape de adevăr.

#### 1.5.4. Sociometria

Sociometria ca metodă a fost folosită pentru prima oară într-un studiu realizat de Moreno în 1934 în Statele Unite în lagăre de persoane deportate și într-o instituție pentru tineri delicvenți. La baza acestei metode există un chestionar ce cuprinde două întrebări care evidențiază alegerea, respectiv respingerea într-un grup în raport cu o anumită activitate. Întrebările pot fi de genul:

- pentru alegere: pe cine ai alege să faci activitatea propusă
- pentru respingere: care ar fi persoana cu care ai dori cel mai mult să nu faci activitatea propusă

Bineînțeles că întrebările pot suferi mici variații, dar astfel încât să se păstreze sensul de alegere, respectiv respingere. În 1952 Tagiuri<sup>29</sup> propune ca la aceste două întrebări să se adauge încă două care să se refere la cine crede cel chestionat că l-ar fi ales și cine crede cel chestionat că l-ar fi respins.

Pentru a avea asigurată validitatea răspunsurilor trebui să respectăm următoarele reguli<sup>30</sup>:

---

<sup>28</sup> În lucrarea [Rad], p. 358.

<sup>29</sup> În lucrarea *Relational analysis: an extension of sociometry method with emphasis upon social perception*.

<sup>30</sup> Cf. [Abr].

- grupul studiat trebuie să aibă o perioadă semnificativ de lungă de conviețuire
- trebuie să avem acordul membrilor grupului, să nu fie impus faptul de a răspunde la întrebări.
- asigurarea confidențialității răspunsurilor
- trebuie să fie explicitate criteriile alegerii
- chestionarul trebuie aplicat la grupuri de maxim 30 de persoane

Indicii ce trebuie urmăriți sunt<sup>31</sup>:

1. numărul de alegeri primite
2. numărul de alegeri făcute
3. numărul de alegeri reciproce
4. numărul de respingeri primite
5. numărul de respingeri făcute
6. numărul de respingeri reciproce

Cercetarea poate fi extinsă prin studierea celor șase indici pentru realizarea unei activități profesionale, cât și pentru realizarea unei acțiuni în timpul liber. Centralizarea datelor se realizează într-o sociomatrice, iar ca reprezentare grafică putem folosi o sociogramă. Modul în care realizăm o sociogramă este descris în capitolul 2 secțiunea 2.3.4.

## 1.6. Măsurarea datelor

La începutul acestui capitol spuneam că ceea ce se obține prin diferite metode de investigare (o parte din ele le-am prezentat în secțiunile anterioare) sunt date. Un prim pas spre transformarea acestor date în informații este măsurarea lor.

În continuare, vom considera datele obținute în urma măsurării, ***date brute***, celelalte tipuri de date numindu-le ***date prelucrate***. Nu am folosit un antonim al expresiei date brute, deoarece prelucrarea datelor nu este niciodată definitivă, orice tip de date pretându-se la prelucrări succesive, în funcție de fenomenul studiat.

---

<sup>31</sup> Cf. [Abr].

Măsurarea este o „operație prin care se atribuie numere datelor discrete sau continue ce urmează a fi evaluate”<sup>32</sup>. Putem realiza măsurarea în două moduri:

- citirea directă a numerelor afișate de aparatele cu care se efectuează experimentul
- prin calcularea scorurilor sau a frecvențelor rezultate în urma aplicării unui instrument specific (test, chestionar etc.)

Din prima categorie amintim:

- experiențele din laboratorul de psihologie (cum ar fi frecvența apariției unei reacții în funcție de un stimul studiat, parametri fizici – greutate, intensitate etc.)
- timpul necesar pentru a răspunde unui stimul

Din a doua categorie amintim:

- aplicarea unui chestionar
- aplicarea unui test (însurarea punctelor conform cu scalele)

În cele ce urmează ne vom referi la măsurătorile din cea de-a doua categorie. Dacă, de exemplu, în fizică există mărimi ce au unități de măsură bine definite (kg pentru masă, newton pentru greutate, m/s pentru viteză etc.), în studiul fenomenelor sociale acest gen de unități de măsură nu există. De aceea, s-a recurs la utilizarea de măsurători prin însumare pentru fenomenele sociale. Dintre nivelurile la care se pot realiza tipuri de măsurători amintim:

- nivelul nominal (calitativ)
- nivelul ordinal
- nivelul hiperordinal<sup>33</sup>

*Măsurători la nivel nominal* reprezintă tipul cel mai des întâlnit în fenomenele sociale. A folosi o măsurătoare bazată pe nivelul nominal înseamnă a ordona elementele măsurate după diferite categorii. Aceste categorii, pentru a putea avea un proces corect de

---

<sup>32</sup> Cf [Rad], p. 47.

<sup>33</sup> Termen preluat din [Rad].

ordonare<sup>34</sup>, trebuie să fie disjuncte, adică un element al unei categorii nu poate aparține, în același timp, și altei categorii.

### **Exemplul 20**

Unui grup *i* se aplică un stimul<sup>35</sup>. În acest caz, putem stabili patru categorii:

- a. nicio reacție
- b. doar reacție nonverbală
- c. doar reacție verbală
- d. și reacție verbală și reacție non-verbală

Categoria *d* a fost pusă deoarece subiecții pot avea reacții și la nivel de gest și la nivel verbal.

### **Exemplul 21**

În cadrul unui chestionar, aplicat unor tineri între 15 și 18 de ani, s-a pus următoarea întrebare deschisă: *Ce înseamnă pentru tine a merge la școală?* După ce s-au citit chestionarele primite, s-au stabilit următoarele categorii:

- a. a învăța
- b. a mă plictisi
- c. conflicte cu ceilalți
- d. pierdere de timp

A mai fost folosită și o a cincea categorie, *e. altele*, pentru acele situații care nu se încadrau în cele patru categorii stabilite, dar nici nu se regăseau în număr destul de mare pentru a constitui o categorie de sine stătătoare.

*Măsurători la nivel ordinal:* acest tip de măsurători este cel mai folosit deoarece, în urma folosirii lui, se pot realiza comparații între subiecți, sau, altfel spus, se poate stabili o relație de ordine între participanții la experiment.

---

<sup>34</sup> Prin ordonare aici înțelegem ordine nu și comparație.

<sup>35</sup> Prin stimul înțelegem chiar și aplicarea unei întrebări.

## Exemplul 22

S-a aplicat testul Hamilton pe un lot de 20 de bărbați și s-au obținut următoarele rezultate:

|     | Hamilton |
|-----|----------|
| S1  | 18       |
| S2  | 16       |
| S3  | 14       |
| S4  | 15       |
| S5  | 20       |
| S6  | 10       |
| S7  | 9        |
| S8  | 14       |
| S9  | 19       |
| S10 | 18       |
| S11 | 15       |
| S12 | 11       |
| S13 | 7        |
| S14 | 20       |
| S15 | 22       |
| S16 | 15       |
| S17 | 10       |
| S18 | 17       |
| S19 | 13       |
| S20 | 17       |

Se observă că, în urma măsurării, s-a calculat pentru fiecare subiect valoarea obținută la test. Pornind de la aceste valori putem compara subiecții din punctul de vedere al testului Hamilton. De exemplu, putem spune că subiectul S4 a obținut un scor mai mare decât subiectul S7.

*Măsurători la nivelul hiperordinal:* La acest gen de măsurători adăugăm, pe lângă nivelul ordinal, și o analiză a distanței dintre elementele situate pe scara de valori. Deși nu foarte folosit,

acest nivel de măsurare reprezintă un prim pas spre utilizarea mulțimilor fuzzy în analiza datelor obținute în sociologie sau psihologie.

**Exemplul 23**

În cadrul unui studiu asupra unui grup de 10 elevi s-a alcătuit o listă cu mediile acestora<sup>36</sup> (datele au fost ordonate descrescător):

|         | Nota |
|---------|------|
| Elev 1  | 10   |
| Elev 8  | 10   |
| Elev 9  | 10   |
| Elev 6  | 7    |
| Elev 10 | 7    |
| Elev 4  | 6    |
| Elev 2  | 5    |
| Elev 3  | 5    |
| Elev 5  | 5    |
| Elev 7  | 5    |

Se observă, din analiza distanțelor dintre notele elevilor, că se pot determina trei grupuri de elevi: primul grup ar putea fi alcătuit din cei care au obținut nota 10 (elevul 1, elevul 8, elevul 9), al doilea grup (din elevii care au notele 6, 7 – elevul 6, elevul 10 și elevul 4) și un al treilea grup (format din elevii ce au nota 5 – elevul 2, elevul 3, elevul 5 și elevul 7). Aceste scări de intervale pot constitui un punct de plecare în analiza datelor obținute din acest experiment.

---

<sup>36</sup> Pentru claritatea expunerii prezentăm exemplul doar pe 10 elevi, în practică însă nu se recomandă să se tragă concluzii pe un grup atât de mic.

## 2. REPREZENTAREA DATELOR CA MOD DE EXAMINARE PRELIMINARĂ A FENOMENULUI STUDIAT

*The preliminary examination of most data is facilitated by the use of diagrams. Diagrams prove nothing, but bring outstanding features readily to the eye; they are therefore no substitute for such critical tests as may be applied to the data, but are valuable in suggesting such tests, and in explaining the conclusions founded upon them*<sup>37</sup>

R.A. Fisher

A prezenta datele nu înseamnă doar a le înșirui pe o foaie de hârtie. Reprezentarea datelor reprezintă o etapă importantă a prelucrării datelor. Înainte de arăta cum putem realiza această stocare a datelor vom prezenta câte ceva despre datele culese din teren, apoi vom arăta diferite moduri de a prezenta datele: prin diagrame, prin scheme, sociograme. O secțiune aparte o va constitui prezentarea *graficului Gantt* ca mod de planificare a unui proiect.

### 2.1. Serii de numere

În această secțiune ne vom referi la seriile de numere așa cum pot fi ele întâlnite în majoritatea proceselor sociale, fără a face referire la semnificația matematică a conceptului de serie, semnificație care este mult mai largă. În acest context vom înțelege prin *serie de numere*, mulțimea de numere ce au o caracteristică comună. Pentru a înțelege mai bine această noțiune vom da două exemple.

---

<sup>37</sup> Pentru traducere vezi finalul acestui capitolul.

### Exemplul 24

Avem un lot de 15 subiecți de sex masculin cărora le aplicăm un test de inteligență. Mulțimea alcătuită din rezultatele la acest test reprezintă o serie de numere, caracteristica lor comună fiind faptul că ne arată coeficientul de inteligență al subiecților din lot. Putem avea o a doua serie de numere ce reprezintă coeficientul de inteligență al subiecților de sex feminin dintr-un lot de 15 subiecți. Problema ce ar putea apărea este de a compara cele două loturi prin prisma acestor două serii de numere.

### Exemplul 25

Să presupunem că avem o întrebare dintr-un chestionar ce este aplicat pe două grupuri de subiecți, unul în mediul rural, iar al doilea în mediul urban. Să presupunem că în urma prelucrării datelor am obținut următoarele rezultate:

| <i>Mediul</i> | <i>Da</i> | <i>Nu</i> |
|---------------|-----------|-----------|
| Rural         | 64%       | 36%       |
| Urban         | 52%       | 48%       |

În acest caz o primă serie de numere o reprezintă procente la răspunsuri date de subiecții din mediul rural, iar o a doua serie o reprezintă procente date de subiecții din mediul urban.

Dacă o serie reprezintă un proces ce se petrece în timp, atunci aceasta se va numi *serie temporală*<sup>38</sup>.

---

<sup>38</sup> Sau serie cronologică

### Exemplu 26

În tabelul de mai jos am măsurat creșterile anuale ale populației globului în perioada 1995-2000

| Anii | Nr. locuitori |
|------|---------------|
| 1995 | 80,781,974    |
| 1996 | 79,253,622    |
| 1997 | 79,551,074    |
| 1998 | 78,019,039    |
| 1999 | 76,861,716    |
| 2000 | 75,529,866    |

Seria este temporală întrucât ne arată un proces – creșterea populației, și acest proces este studiat într-o perioadă de timp – din anul 1995 până în anul 2000.

Analog putem avea *serii de tip spațial*<sup>39</sup>.

### Exemplul 27

În mai multe licee din București se pune următoare întrebare profesorilor: *Se poate vorbi de o „dependență de computer” în rândul elevilor?* În urma prelucrării datelor s-au obținut următoarele rezultate<sup>40</sup>:

| Locul aplicării | Procent răspunsuri afirmative |
|-----------------|-------------------------------|
| Liceu sector 1  | 58 %                          |
| Liceu sector 2  | 49%                           |
| Liceu sector 3  | 61%                           |
| Liceu sector 4  | 41%                           |
| Liceu sector 5  | 40%                           |
| Liceu sector 6  | 44%                           |

În acest caz seria de numere este studiată în aceeași perioadă de timp, dar în puncte diferite din spațiu.

---

<sup>39</sup> Sau *serii teritoriale*.

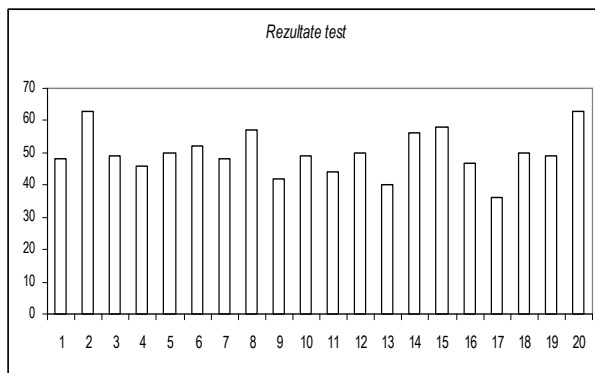
<sup>40</sup> Datele prezentate sunt ipotetice.

## 2.2. Gruparea datelor

În continuare vom da un exemplu pentru a arăta două moduri în care putem reprezenta aceleași date în funcție de modul în care am realizat gruparea.

### **Exemplul 28**

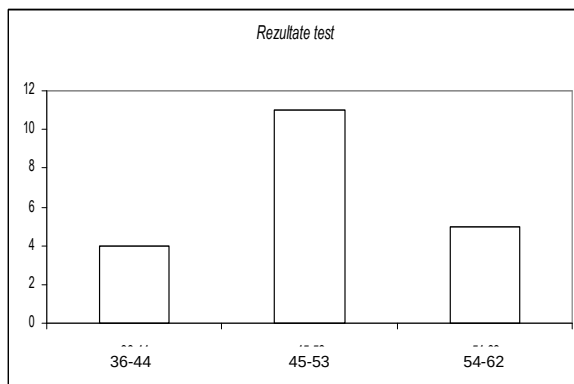
La un cabinet de testare psihologică a șoferilor profesioniști s-a analizat situația subiecților testați în decursul unei săptămâni. Deoarece avem foarte mulți subiecți, realizarea unei diagrame în care să reprezentăm nota obținută de fiecare subiect devine anevoioasă, aceasta nu are nicio relevanță, așa cum se poate observa și din diagrama de mai jos:



În această situație se poate recurge la gruparea datelor pe intervale. Dacă avem lungimea intervalului de 9 puncte atunci datele se vor putea pune astfel într-un tabel:

|                     |       |       |       |
|---------------------|-------|-------|-------|
| <i>intervalul</i>   | 36-44 | 45-53 | 54-62 |
| <i>nr. subiecți</i> | 4     | 11    | 5     |

Astfel am obținut următoarea reprezentare a datelor:



### 2.3. Reprezentări grafice ale seriilor de numere

Când ne referim la reprezentare grafică, înțelegem reprezentarea datelor într-un sistem de coordonate. Reprezentările grafice cu ajutorul unui sistem de coordonate se numesc **diagrame**. Uneori însă putem folosi alte reprezentări cum ar fi: tabele, figuri geometrice, hărți etc.

O diagramă conține, indiferent de formă, un sistem de coordonate și o legendă explicativă atunci când reprezentăm mai multe serii de numere. Diagramele cele mai folosite sunt cele bidimensionale. Așa cum se poate deduce chiar din denumire, diagramele bidimensionale au două axe. Axa  $Ox$ <sup>41</sup> este axa pe care reprezentăm etichetele seriei de numere<sup>42</sup>, iar pe axa  $Oy$ <sup>43</sup> valorile corespunzătoare acestor etichete.

---

<sup>41</sup> Orizontală.

<sup>42</sup> Aceste etichete pot fi sub forma de text sau numere.

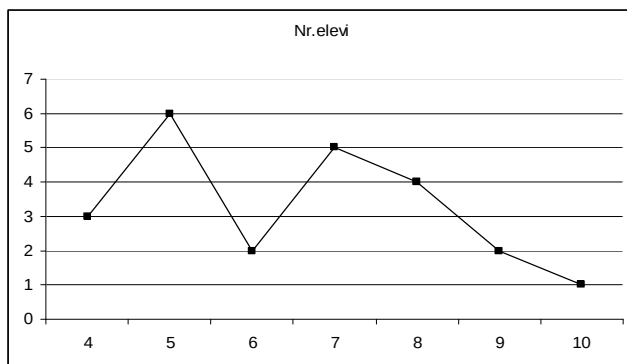
<sup>43</sup> Verticală.

### 2.3.1. Diagramele de tip linie

Reprezentarea datelor prin diagramele de tip linie<sup>44</sup> este unul din modurile cel mai des folosite.

#### **Exemplul 29**

La un test, în urma corectării rezultatelor, s-au obținut următoarele note: cu nota 4 – 3 elevi, cu nota 5 – 5 elevi, cu nota 6 – 2 elevi, cu nota 7 – 4 elevi, cu nota 8 – 3 elevi, cu nota 9 – 4 elevi, cu nota 10 – 3 elevi. Cu rezultatele expuse mai sus realizăm un grafic:



Folosind reprezentarea datelor cu ajutorul graficului, informația obținută devine mai amplă decât în cazul prezentării prin tabel. Putem observa imediat, în exemplul 29, faptul că în clasă există cele două tipuri de elevi: cei care se centrează în jurul notei 5 și cei care se centrează în jurul notei 7.

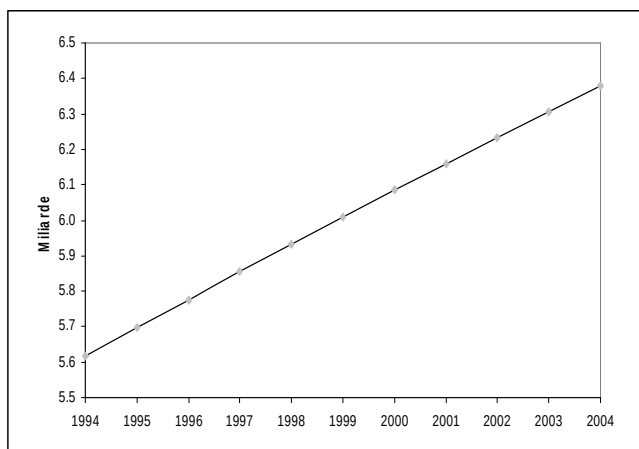
Dacă în loc de note realizăm o analiză a fenomenelor în funcție de timp, atunci graficul ne va arăta evoluția în timp a unui proces:

---

<sup>44</sup> Cunoscută sub denumirea de *grafice*.

### Exemplul 30

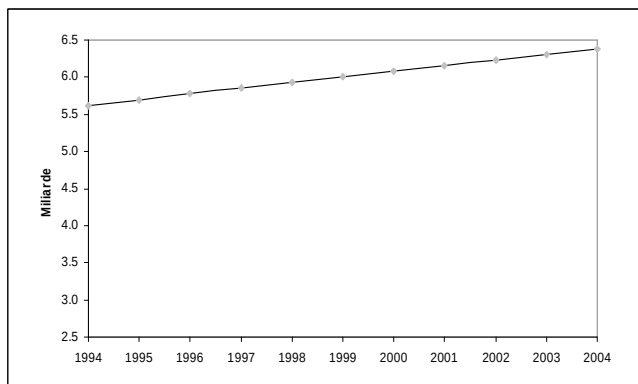
Dorim să analizăm creșterea populației globului în perioada 1994-2004<sup>45</sup>. Atunci vom pune pe axa Ox anii, iar pe Oy valorile. Vom obține următorul grafic:



În cadrul acestui tip de grafic se poate observa faptul că, dacă pe scala Oy schimbăm valorile, forma graficului se schimbă. În cazul de mai sus valorile de pe axa Oy sunt între 5,5 miliarde și 6,5 miliarde. Dacă însă luăm valorile între 2,5 și 6,5 miliarde vom obține următorul grafic:

---

<sup>45</sup> Sursa: *U.S. Census Bureau, International Data Base*. Pentru 2004 este dată desigur o estimare.



În acest caz creșterea nu mai pare atât de pronunțată ca în primul caz. (Iată o primă justificare a faptului că nu ne putem baza pe forma unei diagrame, ci avem nevoie de instrumente obiective, cum ar fi analiza statistică a datelor).

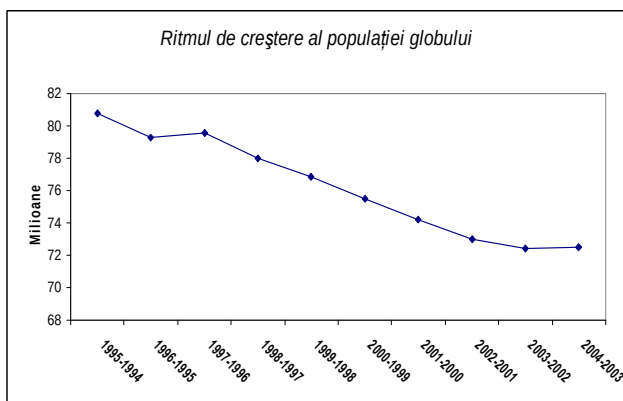
Desigur, apare întrebarea: „Care din cele două diagrame este cea corectă?” Corecte sunt amândouă. Alegerea scalei este dată de problema studiată: Dacă, de exemplu, studiem creșterea populației în raport cu anul 1994, atunci este recomandabilă prima reprezentare. Dacă însă dorim să ne raportăm la anul 1950 când populația globului era de aproximativ 2,5 miliarde atunci este recomandabilă cea de-a doua reprezentare.

### Exemplu 30 (continuare)

Putem reprezenta, tot în aceeași perioadă, cu cât a crescut populația în fiecare an comparativ cu anul anterior. Vom obține datele:

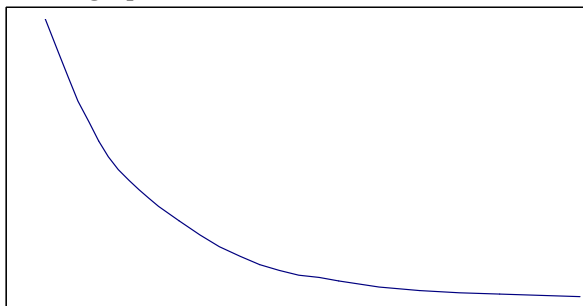
|           |            |
|-----------|------------|
| 1995-1994 | 80,781,974 |
| 1996-1995 | 79,253,622 |
| 1997-1996 | 79,551,074 |
| 1998-1997 | 78,019,039 |
| 1999-1998 | 76,861,716 |
| 2000-1999 | 75,529,866 |
| 2001-2000 | 74,220,528 |
| 2002-2001 | 73,002,863 |
| 2003-2002 | 72,442,511 |
| 2004-2003 | 72,496,962 |

Ca reprezentare vom folosi din nou diagrama de tip linie. După cum se observă din diagrama de mai jos, deși populația globului continuă să crească, se observă o scădere continuă a ritmului de creștere.

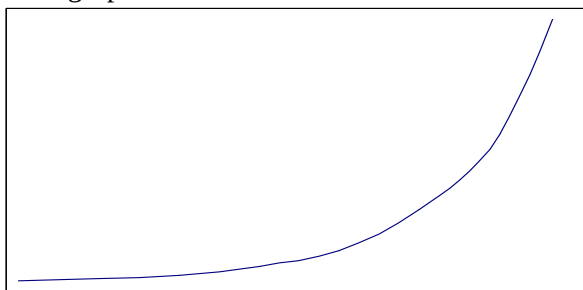


În lucrul cu diagrame de tip linie putem întâlni următoarele tipuri de reprezentări:

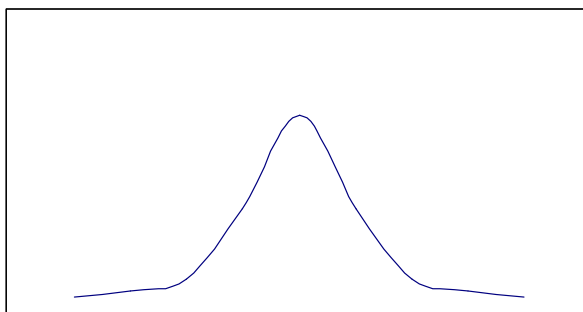
a) datele grupate la cea mai mică valoare<sup>46</sup>



b) date grupate la cea mai mare valoare



c) datele distribuite normal



---

<sup>46</sup> Putem observa chiar că ultimul exemplu se încadrează în această clasă de reprezentări.

În practică, desigur, se pot întâlni mult mai multe tipuri de grafice, dar ne-am oprit la acestea și prin prisma noțiunilor ce vor fi predate în capitolele următoare.

### 2.3.2. *Histogramele*

Un alt mod de reprezentare a datelor este prin histograme. Acest tip de diagramă îl folosim atunci când dorim să comparăm valorile între anumite categorii.

#### ***Exemplul 31***

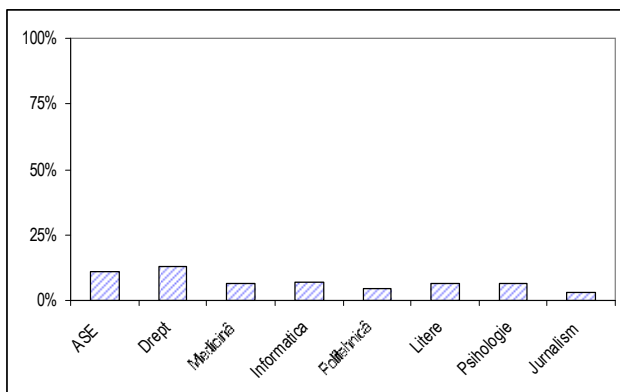
În perioada noiembrie 2003 – februarie 2004, s-a pus următoarea întrebare elevilor de liceu din clasa a 11-a și a 12-a: „La ce facultate veți da examen după bacalaureat?” În urma centralizării datelor s-au obținut următoarele rezultate<sup>47</sup>:

| <b>Facultate</b>   | <b>Procent</b> |
|--------------------|----------------|
| <i>ASE</i>         | 10,7%          |
| <i>Drept</i>       | 12,8%          |
| <i>Medicină</i>    | 6,7%           |
| <i>Informatica</i> | 7,0%           |
| <i>Politehnică</i> | 4,4%           |
| <i>Litere</i>      | 6,6%           |
| <i>Psihologie</i>  | 6,3%           |
| <i>Jurnalism</i>   | 2,8%           |

Ca observație, deși întrebarea se referea clar la facultatea la care vor da examen, totuși se observă că un procent important au dat ca răspuns o instituție întreagă (cum ar fi Academia de Studii Economice). Întrucât acest lucru l-am considerat important pentru studierea opțiunilor pentru o facultate, am lăsat aceste opțiuni.

---

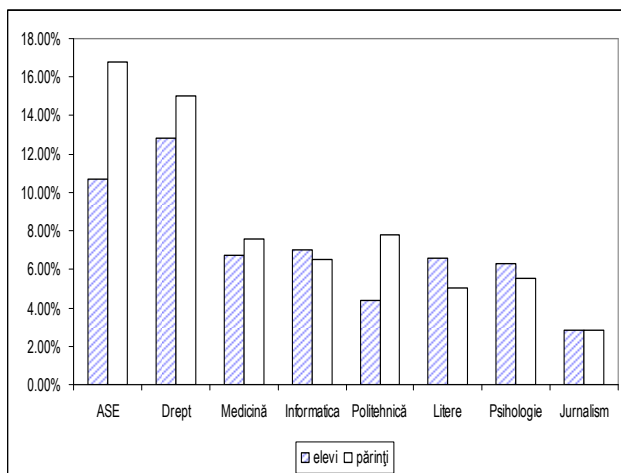
<sup>47</sup> Am pus doar opțiunile ce au întrunit mai mult de 2%.



În acest caz pe axa Ox am pus etichetele (ASE, Drept, Medicină etc.), iar pe axa Oy am pus valorile corespunzătoare fiecărei etichete. Tot în cadrul cercetării respective am pus o întrebare pe aceeași temă și părinților elevilor: „La ce facultate va da examen după bacalaureat fiul/fiica dumneavoastră? ”

| <i>Facultatea</i> | <i>Opțiuni elevi</i> | <i>Opțiuni părinți</i> |
|-------------------|----------------------|------------------------|
| ASE               | 10,7%                | 16,8%                  |
| Drept             | 12,8%                | 15,0%                  |
| Medicină          | 6,7%                 | 7,6%                   |
| Informatica       | 7,0%                 | 6,5%                   |
| Politehnică       | 4,4%                 | 7,8%                   |
| Litere            | 6,6%                 | 5,0%                   |
| Psihologie        | 6,3%                 | 4,1%                   |
| Jurnalism         | 2,8%                 | 2,8%                   |

Am pus pe aceeași diagramă răspunsurile elevilor și ale părinților.



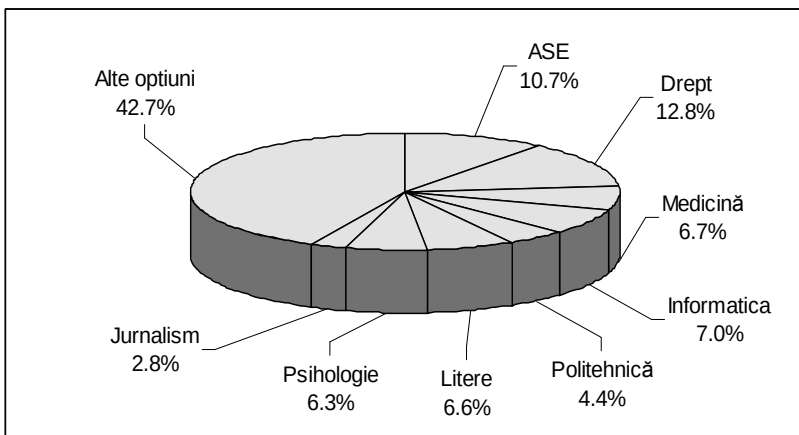
**Observație.** Atunci când am realizat diagrama doar cu opțiunile elevilor am pus pe axa Oy valorile de la 0% la 100%, pentru a putea realiza o primă analiză a opțiunilor elevilor în raport cu totalul opțiunilor elevilor. În al doilea caz, am pus pe axa Oy valori între 0% și 18 % pentru a scoate în evidență eventualele diferențe între opțiunile părinților și ale elevilor.

### 2.3.3. Alte tipuri de diagrame

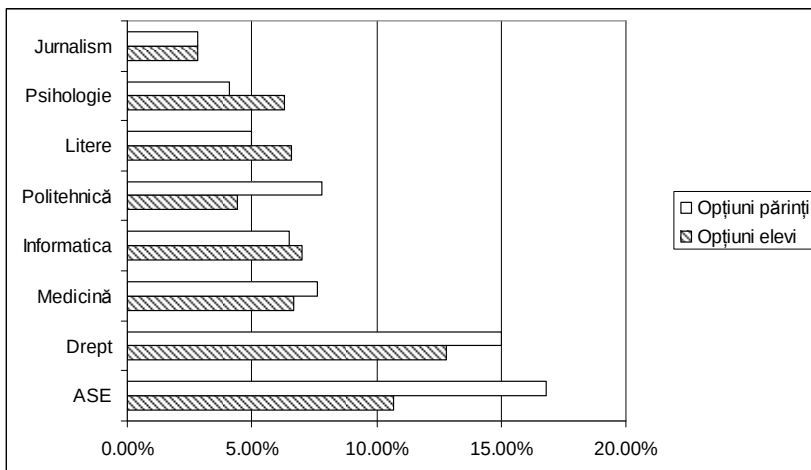
Reprezentarea pe care o putem da datelor se poate realiza în mai multe moduri (datele sunt cele de la exemplul 30).

- *diagrama de tip disc*<sup>48</sup> – dacă dorim să scoatem în evidență procentul opțiunilor elevilor raportându-ne la total. Dar, în această situație, ar trebui să mai introducem și restul opțiunilor sau, pentru simplificarea reprezentării, putem să mai introducem o nouă categorie, *alte opțiuni*. În acest caz obținem următoarea diagramă:

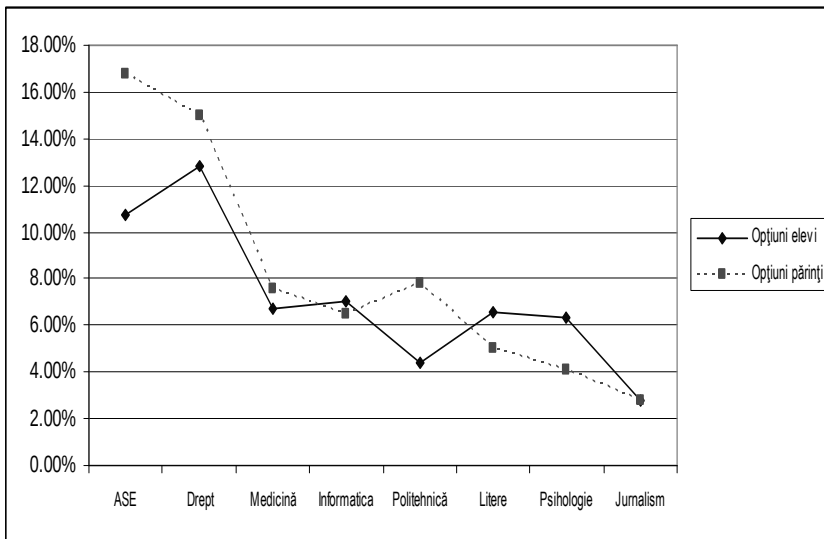
<sup>48</sup> În engleză – Pie.



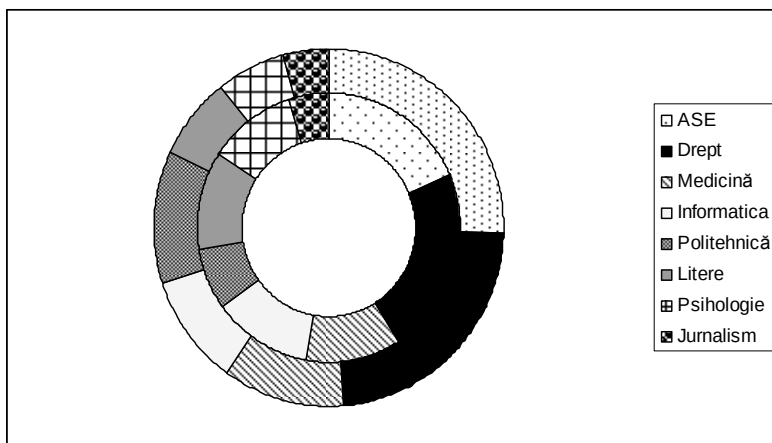
➤ *diagrama de tip bar*



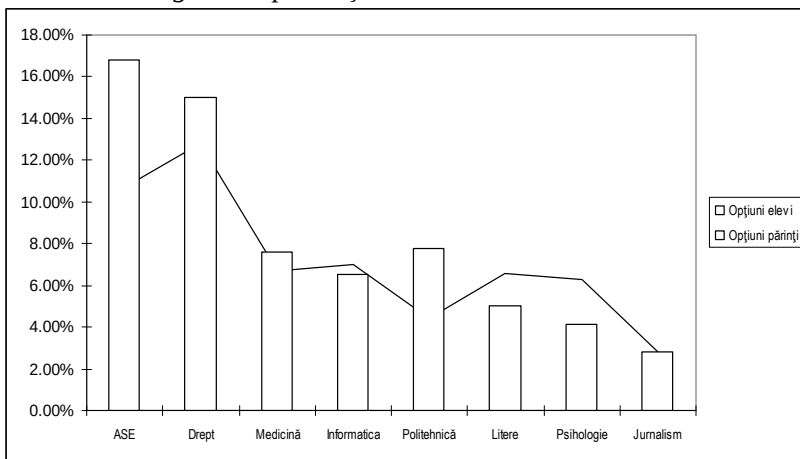
➤ *diagrama de tip linie*



➤ *diagrama de tip inel (doughnut)*

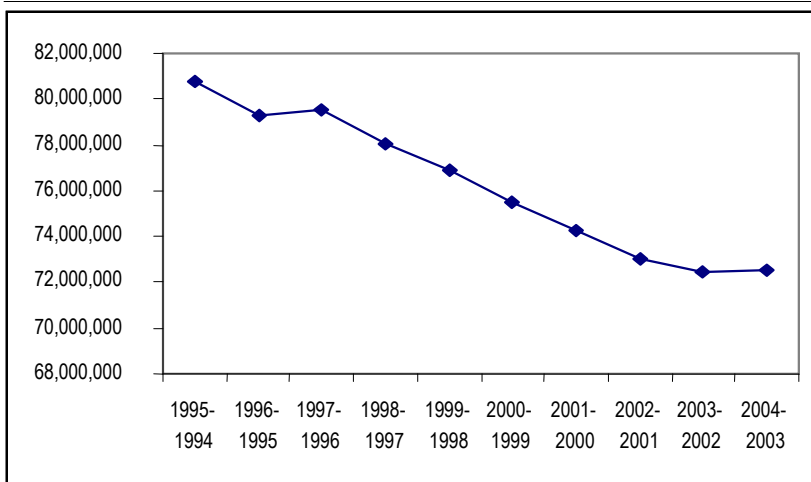


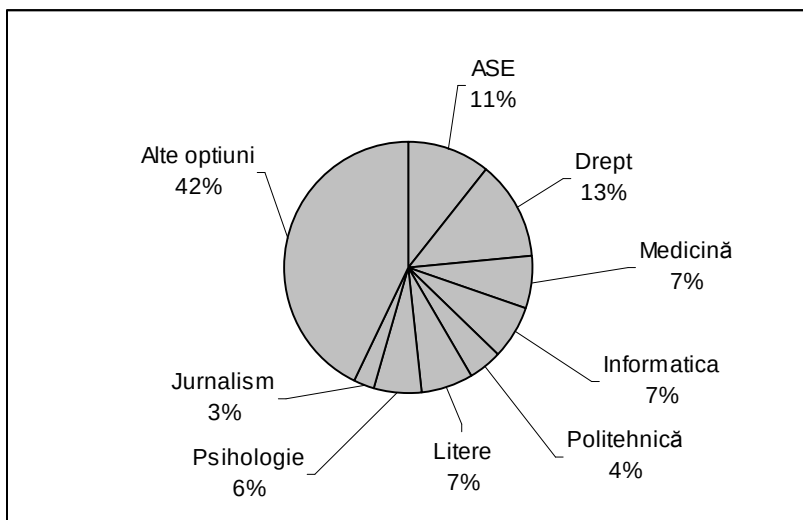
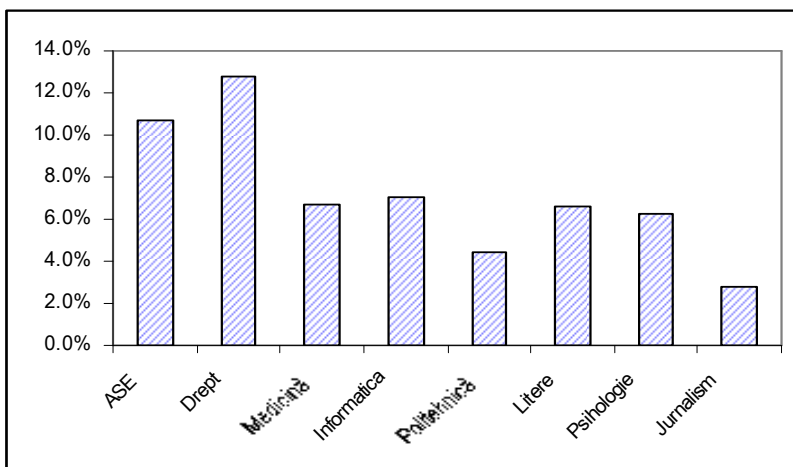
➤ *diagrama tip arie și coloane*



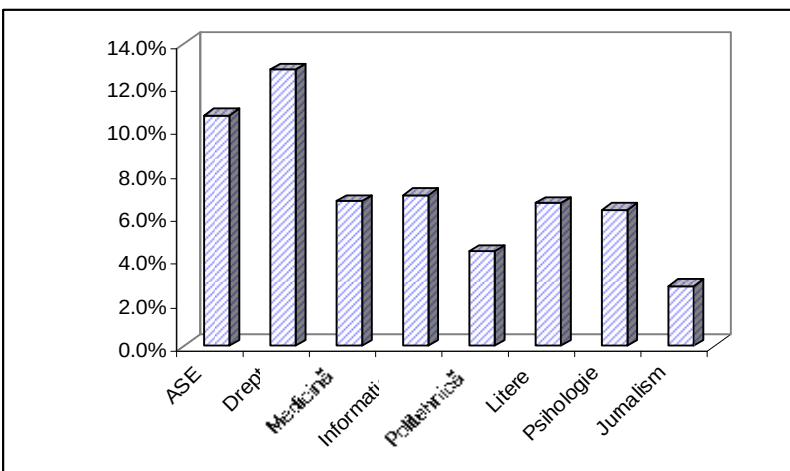
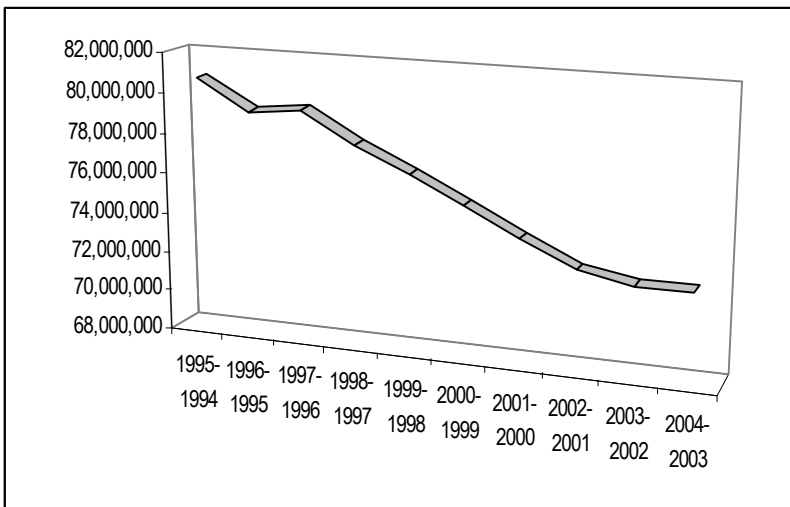
Un alt mod de a clasifica diagramele este raportarea la numărul de coordonate în care este realizată diagrama. Cele mai uzuale sunt: bidimensional și tridimensional.

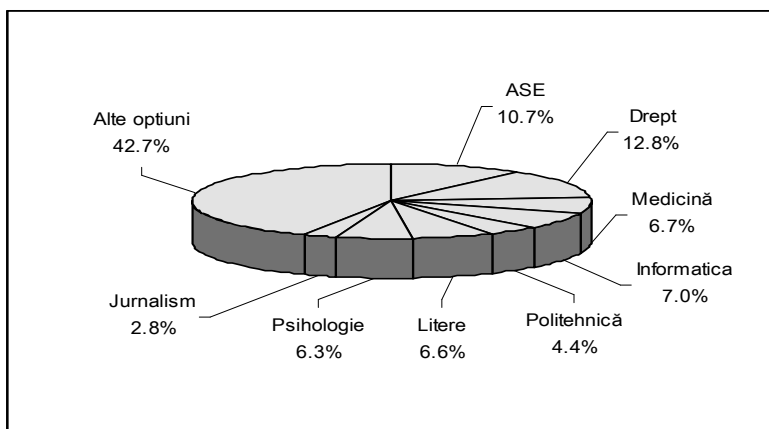
### ***Bidimensional***





## *Tridimensional*





#### 2.3.4. Sociograme

O primă etapă în aplicarea sociometriei este realizarea sociomatricei.

#### Exemplul 32

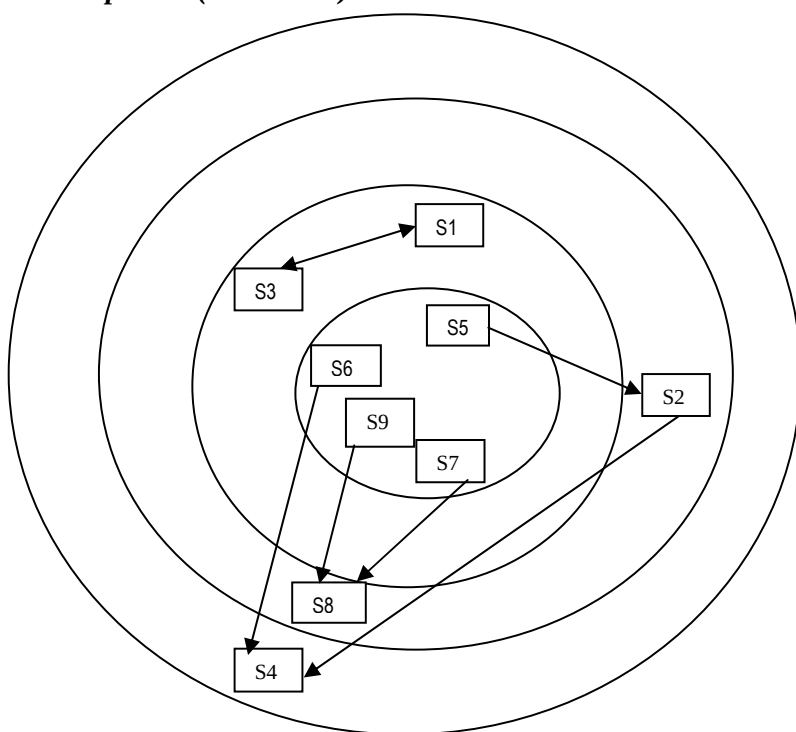
Într-o echipă de lucru s-a pus problema realizării de subechipe de câte doi muncitori care să execute anumite operațiuni împreună și a găsirii unui coordonator al acestor subechipe. Pentru aceasta s-au pus o serie de întrebări membrilor echipei. Una din întrebări a fost: „Dacă ar trebui să faci o deplasare în vederea executării unei lucrări de o zi pe un alt șantier cu cine ai merge?” Subiecții au dat următoarele răspunsuri<sup>49</sup>:

|               |    | cei care sunt aleși |    |    |    |    |    |    |    |    |
|---------------|----|---------------------|----|----|----|----|----|----|----|----|
|               |    | S1                  | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 |
| cei care aleg | S1 |                     |    | *  |    |    |    |    |    |    |
|               | S2 |                     |    |    | *  |    |    |    |    |    |
|               | S3 | *                   |    |    |    |    |    |    |    |    |
|               | S4 |                     | *  |    |    |    |    |    |    |    |
|               | S5 |                     | *  |    |    |    |    |    |    |    |
|               | S6 |                     |    |    | *  |    |    |    |    |    |
|               | S7 |                     |    |    |    |    |    |    | *  |    |
|               | S8 |                     |    |    | *  |    |    |    |    |    |
|               | S9 |                     |    |    |    |    |    |    | *  |    |

<sup>49</sup> Le-am sistematizat sub forma unui tabel.

Putem observa că S5, S6, S7, S9 nu au nicio opțiune, că S1 și S3 au o opțiune, S2 și S8 au două opțiuni și S4 are trei opțiuni. Este însă destul de greu ca din tabelul de mai sus să punem în evidență și relațiile între subiecți. Un mod de reprezentare ar putea fi sub formă de cercuri concentrice, unde în cercul interior vom pune pe cei ce au 0 opțiuni, iar în cercul cel mai mare vom pune pe cei cu 3 opțiuni. În plus, vom duce niște săgeți de la subiectul ce a ales la subiectul ales. Dacă doi subiecți s-au ales între ei, de exemplu S1 cu S3, atunci săgeata va avea dublu sens.

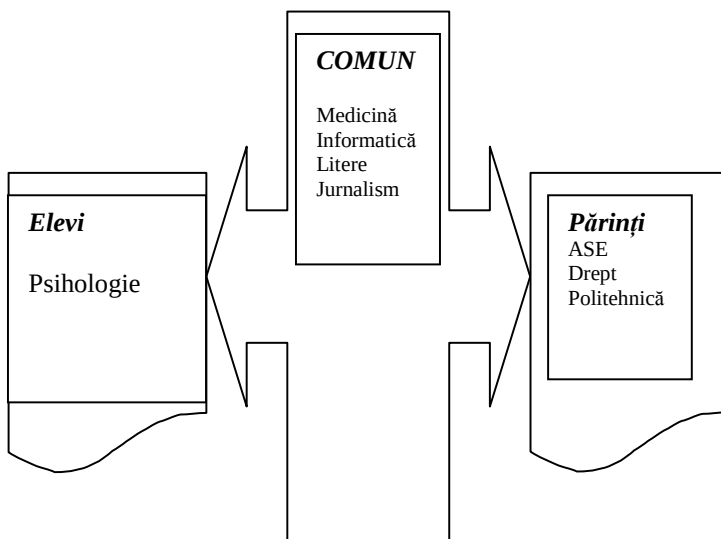
### ***Exemplul 32 (continuare)***



În capitolul 4, în secțiunea 4.1.6. vom vedea că putem continua prelucrarea acestor date prin realizarea grafului asociat și a unei matrice asociate sociogramei.

### 2.3.5. Alte reprezentări ale datelor

Așa cum am specificat la începutul acestui capitol, în afară de diagrame, mai există o serie de reprezentări pentru analiza datelor. Pentru a explica mai bine vom reveni la exemplul 30. Să presupunem că dorim să analizăm comparativ opțiunile părinților cu cele ale elevilor. Pentru aceasta putem folosi următoarea regulă: dacă diferențele procentuale între opțiunile părinților și opțiunile elevilor sunt mai mici de 2% atunci vom spune că acestea nu diferă. În această situație vom avea opțiuni comune pentru părinți și elevi și opțiuni specifice fiecărei categorii. Putem reprezenta aceste date astfel:



Desigur, pentru figura de mai sus se poate folosi orice tip de formă.

## 2.4. Noțiuni introductive de prezentare a datelor în Power Point

Odată ce am găsit modul de reprezentare (diagrame, scheme etc.) a datelor, atunci când ne gândim la prezentarea lor în public (fie că e vorba de comisia de licență, fie la o sesiune de comunicări), trebuie să găsim un mod de prezentare cât mai clar și ușor de folosit. O primă metodă este realizarea unui poster. Ce este un poster? În general e o coală de carton cam de un metru înălțime și 50 de cm lungime pe care se lipesc patru-cinci foi format A4 cu diagrame și/sau diferite tabele cu date mai importante. La sesiunile de comunicări se alocă timp distinct pentru acest gen de prezentări unde autorul (autorii) pun (e) posterul în zona special amenajată pentru acest gen de prezentări și răspund la eventualele întrebări puse de cei ce citesc și sunt interesați de ceea ce se prezintă în poster.

O altă metodă este de a pregăti niște transparente ce se pot pune la un retroproiector, fiecare folie transparentă fiind pusă de cel ce prezintă și dă explicații audiențelor.

O a treia metodă este de a pregăti pe un computer toată prezentarea și de a prezenta folosind un videoproiector.

Desigur, mai pot fi și alte metode de prezentare, dar ne-am mărginit la acestea trei întrucât reprezintă cele mai răspândite moduri de prezentare care se pot folosi în toate universitățile.

Materialul pentru toate cele trei metode poate fi pregătit bazându-ne pe un soft specializat: Power Point, Open Office, Staroffice etc. Deoarece la ora actuală cel mai cunoscut este Power Point vom face o scurtă prezentare a acestuia.

### 2.4.1. Ce este Power Point?

Power Point reprezintă un soft specializat pentru prezentări profesionale. Concret, el are multe din facilitățile celorlalte softuri de la Microsoft, cum ar fi: editare text, tabelare, realizarea de desene, importarea de fișiere din alte programe. Când se creează o prezentare cu Power Point, aceasta se face pe bază de *slide*—uri<sup>50</sup>. Prezentarea se poate realiza într-un singur fișier și la ea, pe lângă imagine, se pot atașa și

---

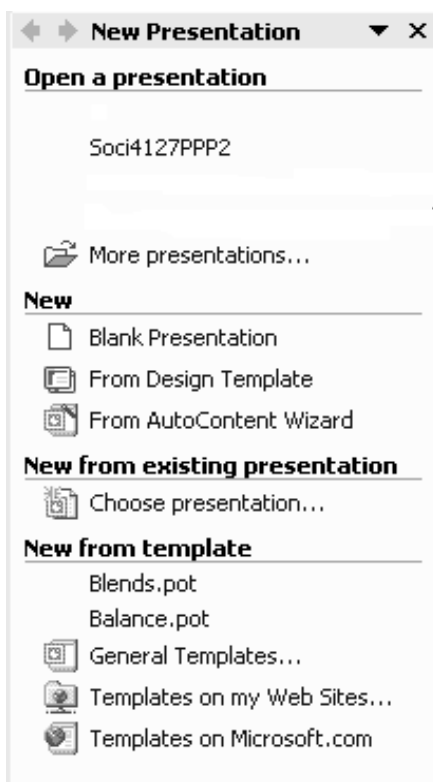
<sup>50</sup> Vom păstra denumirea în engleză deoarece în dicționar la cuvântul *slide* – am găsit *diapozitiv*, *lamelă* ceea ce mi s-a părut că ar reduce înțelegerea posibilităților acestui instrument.

fișiere audio (cum ar fi vocea prezentatorului sincronizată cu parcurgerea automată a slide-urilor sau o melodie etc.)

#### 2.4.2. Cum începem?

S-ar putea ca icon-ul să fie pe ecran (desktop) și atunci, dând un dublu-clic pe acest icon, se deschide Power Point. Dacă nu vedeți icon-ul pe ecran atunci dați un clic pe butonul de START (se află de obicei în stânga-jos pe ecran) și din fereastra ce se deschide alegeți *Programs*, apoi *Power Point*.

Pe ecran, în partea dreaptă va apărea o fereastră ca cea de mai jos:



*Open a presentation* – așa cum arată și titlul, în această parte a ferestrei se găsește o listă cu ultimele fișiere salvate. Dacă dorim să deschidem unul din acest fișiere este suficient să dăm un clic pe denumirea fișierului.

*New* – folosind această parte a ferestrei putem crea o prezentare slide cu slide pornind de la slide-uri albe<sup>51</sup> (*Blank Presentation*), să realizăm o prezentare pornind de la slide-uri preformate (*From Design Template*) sau să creăm o prezentare fiind ghidați de un program specializat (*From AutoContent Wizard*)

*New from existing presentation* – crearea unei noi prezentări pornind de la o prezentare deja salvată pe computer.

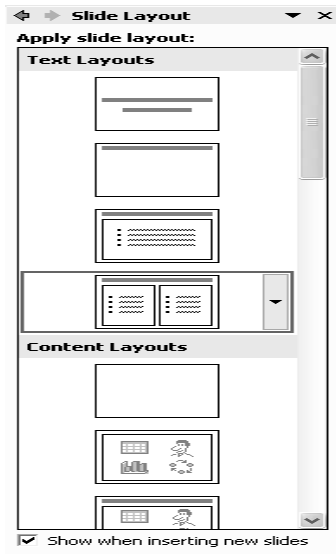
*New from templates* – realizarea unei noi prezentări pornind de la forme predefinite fie de pe computer, fie de pe WEB

**Observație.** Diferența dintre *From Design Template* și *New from templates* constă în faptul că, în timp ce, în primul caz, afișarea slide-urilor preformate se face în fereastra din stânga ecranului, în al doilea caz slide-ul preformat se alege dintr-o listă de slide-uri.

Acum să alegem opțiunea *Blank Presentation* și să dăm un clic pe aceasta. Va apărea pe ecran în centru o foaie albă cu două dreptunghiuri tra-sate cu linie punctată. În cel de sus putem să introducem titlul slide-ului, iar în cel de jos un text. Tot în momentul în care am dat clic pe opțiunea *Blank Presentation* în partea dreaptă a ecranului va apărea următoarea fereastră:

---

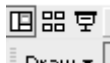
<sup>51</sup> Mi s-a părut mai aproape de ceea ce este afișat denumirea de *slide alb* decât cea de *slide gol*.



Dacă dorim, de exemplu, ca titlul sa fie mai sus, iar sub acest titlu să avem două coloane, dăm un clic pe icon-ul încadrat. Textul, inserarea de imagini etc. se realizează ca într-un editor de texte. În continuare, vom arăta modul în care se poate realiza vizualizarea acestor slide-uri.

#### 2.4.3. Cum se realizează vizualizarea în Power Point?

În fereastra de lucru, în partea stânga-jos avem un grup de trei butoane:



Le vom prezenta pe rând:

Cel mai din stânga buton  este butonul activat implicit și vizualizează câte un slide, astfel încât se pot face modificări pe acel slide afișat. În acest mod avem acces la toate opțiunile din meniul Power Point.

Următorul buton  vizualizează în modul Outline, adică putem vedea slide-urile realizate în ordinea lor, nu putem face modificări în slide-uri, dar putem schimba ordinea lor astfel: mergem cu mouse-ul pe slide-ul căruia dorim să-i schimbăm poziția, când ajunge cursorul pe acel slide, ținem apăsat butonul din stânga al mouse-ului și apoi mișcăm mouse-ul până aducem slide-ul în poziția dorită. Atunci ridicăm degetul de pe mouse.

Cel mai din dreapta buton , numit *slide-show*, este butonul prin care putem intra în modul prezentare a slide-urilor pornind de la slide-ul afișat inițial. În secțiunea ce urmează vom prezenta acest mod de vizualizare.

#### 2.4.4. Modul de vizualizare Slide-Show

Când se folosește acest mod de vizualizare, slide-ul va ocupa tot ecranul, astfel încât de pe ecran vor dispărea bara de titlu, meniurile și orice alt element ce nu face parte din slide. În acest moment computerul e folosit ca un proiector de slide-uri <sup>52</sup>. Atunci când utilizatorul își pregătește prezentarea, are pentru acest mod o serie de opțiuni privind modul în care să apară titlurile pe pagină (simularea „veniri” titlurilor din diferite părți ale ecranului, rotirea titlului până ajunge în poziția destinată etc.), a vitezei de succesiune a slide-urilor etc.

Desigur, în aceste câteva pagini nu ne-am propus decât să facem o scurtă prezentare a posibilităților de prezentare a soft-ului Power Point. Cei ce vor să afle mai multe despre acest mod computerizat de prezentare a datelor îi sfătuim să meargă la manualele de specialitate cum ar fi [Bott], [Kin].

### 2.5. Graficul Gantt

Graficul Gantt reprezintă un instrument de planificare care poate fi folosit pentru alocarea timpului pentru diferite etape ce trebuie realizate în cadrul unui proiect. Acest tip de reprezentare este printre cele mai folosite de către manageri în planificarea diferitelor proiecte deoarece este ușor de înțeles și de construit.

Fiecare sarcină este așezată pe câte o linie. Pe coloane se pun datele (zile, săptămâni sau luni – în funcție de durata proiectului). Sarcinile pot decurge secvențial sau în paralel.

---

<sup>52</sup> Montat la un video proiector poate transmite pe un ecran imaginile de pe slide-uri.

În realizarea unui grafic Gantt trebuie să ținem seama de faptul că avem un număr rezonabil de sarcini astfel încât să le putem gestiona.

### Exemplul 33

Să vedem cum s-ar putea pune, sub forma unui grafic Gantt, etapele realizării unui proiect de diplomă. Să presupunem că avem următoarele etape de realizare a unui proiect:

| Nr | Denumirea activității                                   |
|----|---------------------------------------------------------|
| 1  | Stabilirea obiectivelor                                 |
| 2  | Determinarea ipotezelor ce rezultă din fiecare obiectiv |
| 3  | Stabilirea lotului de lucru                             |
| 4  | Stabilirea instrumentelor                               |
| 5  | Culegerea datelor de pe teren                           |
| 6  | Centralizarea, prelucrarea și analizarea datelor        |
| 7  | Revenirea pe teren pentru a mai culege date de pe teren |
| 8  | Centralizarea, prelucrarea și analizarea noilor date    |
| 9  | Finalizarea lucrării                                    |

Atunci graficul Gantt este:

| Activitatea                                             | nr săptămâni |   |   |   |   |   |   |   |   |    |    |    |    |    |    |
|---------------------------------------------------------|--------------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
|                                                         | 1            | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Stabilirea obiectivelor                                 | ■            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |
| Determinarea ipotezelor ce rezultă din fiecare obiectiv |              |   |   |   |   |   |   |   |   |    |    |    |    |    |    |
| Stabilirea lotului de lucru                             | ■            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |
| Stabilirea instrumentelor                               | ■            |   |   |   |   |   |   |   |   |    |    |    |    |    |    |
| Culegerea datelor de pe teren                           |              | ■ | ■ | ■ |   |   |   |   |   |    |    |    |    |    |    |
| Centralizarea, prelucrarea și analizarea datelor        |              |   |   | ■ | ■ | ■ | ■ | ■ |   |    |    |    |    |    |    |
| Revenirea pe teren pentru a mai culege date de pe teren |              |   |   |   |   |   |   | ■ | ■ | ■  |    |    |    |    |    |
| Centralizarea, prelucrarea și analizarea noilor date    |              |   |   |   |   |   |   |   |   | ■  | ■  | ■  | ■  |    |    |
| Finalizarea lucrării                                    |              |   |   |   |   |   |   |   |   |    |    | ■  | ■  | ■  | ■  |

## 2.6. Concluzii

În finalul acestui capitol dorim să reluăm ideea lui R.A. Fisher în ceea ce privește rolul reprezentării datelor în analiza și interpretarea lor: „Examinarea preliminară a celor mai multe date este facilitată de folosirea diagramelor. Diagramele nu demonstrează nimic, dar aduc caracteristici excelente mai aproape de privire; de aceea ele nu reprezintă un înlocuitor pentru teste ce pot fi aplicate datelor, dar sunt valoroase în analiza acestor teste și în a explica concluziile bazate pe ele”.<sup>53</sup>

<sup>53</sup> Cf. Fisher R.A. (1925) – *Statistical methods for research workers* – Oliver and Boyd, London.

### 3. STATISTICA ÎN CERCETAREA SOCIALĂ

*Celor care spun că se poate demonstra orice cu statistica le răspund: cu statistica prost folosită se poate demonstra orice*

*R.A. Fisher*

În cele ce urmează vom prezenta anumite noțiuni din statistica matematică. În general, prezentarea va fi mai mult intuitivă. Pentru o prezentare mai riguroasă a noțiunilor recomandăm consultarea unui curs de statistică socială<sup>54</sup>.

Fiecare secțiune va avea, la început, exemple de cercetări care reprezintă diferite moduri de a pune problema. Răspunsurile pentru problemele puse la începutul fiecărei secțiuni se vor găsi în secțiunea 3.13.

#### 3.1. Caracteristici (variabile)

##### **Exemplul 34**

Pentru proiectul de diplomă un student și-a ales tema: „Complexul de inferioritate la adolescenți”. Ce ar trebui să cerceteze studentul în cadrul acestui proiect?

##### **Exemplul 35**

La două clase a VI-a din aceeași școală, existau performanțe școlare semnificativ diferite, una din clase având rezultate foarte bune și foarte mulți copii participanți la olimpiade cu rezultate foarte bune, în timp ce, la cealaltă clasă, deși aveau cam aceiași profesori, problema principală a elevilor era să nu rămână corigenți. Pe ce ar trebui să se axeze un eventual studiu de determinare a cauzelor diferențelor între performanțele școlare ale elevilor?

---

<sup>54</sup> De exemplu, [Alex].

Ceea ce face ca anumite elemente să fie grupate pentru analiză constituie o caracteristică comună. Numerele rezultate, în urma aplicării unui test de inteligență, vor putea fi analizate împreună deoarece au comună caracteristica dată de coeficientul de inteligență. Caracteristicile poartă numele de *variabile*. Deci, pentru a defini o variabilă, avem nevoie de o etichetă pentru elementul măsurat și acestei etichete i se asociază o proprietate<sup>55</sup>.

### Exemplul 36

Pentru subiecții cărora li s-a aplicat un test de inteligență se poate reprezenta variabila sub următoarea formă:

| Etichetă                  | Subiect 1 | Subiect 2 | Subiect 3 | Subiect 4 | Subiect 5 |
|---------------------------|-----------|-----------|-----------|-----------|-----------|
| Valoare coef. de intelig. | 117       | 100       | 98        | 79        | 102       |

Întrucât ceea ce ne va interesa este caracterizarea unui grup de elemente, vom folosi, în continuare, denumirea de *caracteristică*.

Caracteristicile vor fi de tip calitativ sau cantitativ. Caracteristicile de tip calitativ, așa cum se poate deduce din denumire, ne arată o calitate a elementelor grupului.

### Exemplu 37

Putem avea caracteristica *grad universitar* și atunci ea ar putea fi de forma:

| Etichetă         | Subiect 1    | Subiect 2 | Subiect 3 | Subiect 4  | Subiect 5 | Subiect 6 |
|------------------|--------------|-----------|-----------|------------|-----------|-----------|
| Grad universitar | Conferențiar | Lector    | Lector    | Preparator | Lector    | Profesor  |

Caracteristicile de tip cantitativ sunt cele în care studiem valori asociate elementelor grupului.

---

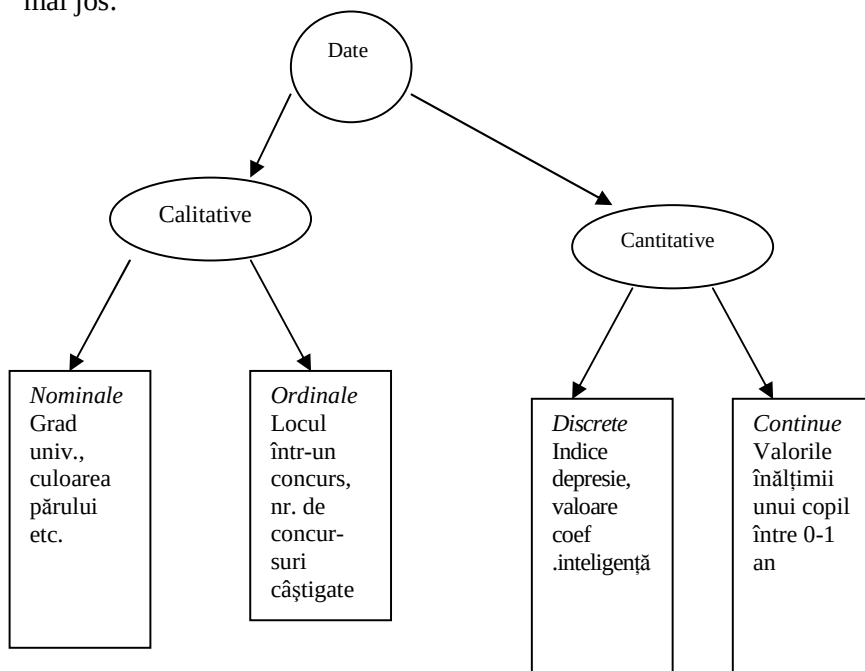
<sup>55</sup> Am folosit termenul de *proprietate* pentru a cuprinde amândouă posibilitățile: Etichetei să-i fie asociată o valoare sau o calitate.

### Exemplul 38

Dacă studiem indicele de depresie la un grup, caracteristica va arăta astfel:

| Etichetă           | Subiect 1 | Subiect 2 | Subiect 3 | Subiect 4 |
|--------------------|-----------|-----------|-----------|-----------|
| Indice de depresie | 55        | 40        | 46,25     | 32,5      |

O abordare mai riguroasă a datelor se poate da prin schema de mai jos:



## 3.2. Populație

### Exemplul 39

Pentru realizarea unui sondaj privitor la alegerile locale dintr-un anumit oraș, din cine se compune mulțimea pe care o studiem? Dar dacă sondajul s-ar referi la alegerile generale?

Populația<sup>56</sup>, din punct de vedere statistic, reprezintă mulțimea de elemente care au una sau mai multe caracteristici comune. De exemplu, putem avea populația alcătuită din persoanele din mediul urban, în acest caz caracteristica comună fiind mediul. Putem avea și o populație alcătuită din persoanele de sex feminin din mediul rural. În această situație, persoanele au două caracteristici comune ce le reunesc într-o populație: faptul că sunt de sex feminin, și că sunt din mediu rural. Elementele populației pot fi: persoane, obiecte, evenimente, idei, opinii etc.

Putem spune că asupra elementelor unei populații acționează aceleași legi statistice din punct de vedere al caracteristicii studiate. În raport cu aceste legi se poate defini gradul de omogenitate al unei populații. Pentru studiile în științele sociale această omogenitate ne poate fi foarte folositoare în realizarea de predicții asupra populației studiate. Un posibil rezultat ar putea fi: cu cât omogenitatea este mai mare cu atât gradul de siguranță al predicției va fi mai mare.

#### **Exemplul 40**

Luând caracteristica – *învățarea limbii franceze la școală* în clasa a 6-a, putem avea o populație alcătuită din elevii de clasa a 6-a ce învață limba franceză la școală. S-ar putea ca omogenitatea acestei populații să fie foarte mică, datorită faptului că elevii provin din mediul urban, iar alții din mediul rural. Desigur, în acest caz, putem transforma studiul și vom spune că analizăm două populații: o primă populație compusă din elevii de clasa a 6-a ce învață franceză și sunt din mediu urban și o a doua populație compusă din elevii de clasa a 6-a ce învață franceză și sunt din mediul rural. Să observăm că fiecare populație are, în fapt, 3 caracteristici: învață limba franceză, sunt în clasa a 6-a și mediul căruia aparțin.

---

<sup>56</sup> O altă denumire pentru populație este și *colectivitate statistică*.

### **Exemplul 41**

Să presupunem că dorim să realizăm un studiu asupra cumpărătorilor dintr-un magazin. În acest caz vom analiza două populații: o primă populație, alcătuită din localnicii ce cumpără din acel magazin, și o a doua populație, alcătuită din turiștii ce au făcut cumpărături în acel magazin. Dacă însă dorim să comparăm cumpărăturile realizate de localnici cu cumpărăturile realizate de turiști, atunci s-ar putea să obținem o omogenitate mai mică a celor 2 populații și atunci s-ar putea să fim nevoiți să studiem după o altă caracteristică, să zicem sexul cumpărătorilor. În acest caz vom avea 4 populații: populația de localnici de sex feminin ce cumpără din magazin, populația de turiști de sex feminin ce cumpără din magazin, populația de localnici de sex masculin ce cumpără din magazin și populația de turiști de sex masculin ce cumpără din magazin. Ca exercițiu, vă puteți imagina, ca o caracteristică de care se poate ține cont în realizarea populației, vârsta celor ce fac cumpărături în acel magazin.

Desigur, împărțirea după prea multe caracteristici duce la un număr mare de populații studiate, ceea ce îngreunează considerabil cercetarea. Putem „uni” două populații într-una singură atunci când diferențele între ele nu sunt semnificative.

La nivel conceptual, totul pare destul de corect, însă din punct de vedere practic, chiar și cazul în care avem de analizat caracteristicile unei singure populații apar destule probleme<sup>57</sup>. Pentru a înțelege mai bine acest lucru este suficient să ne gândim la recensământul populației unei țări. Am dat acest exemplu pentru a arăta că totuși analiza fiecărui individ al unei populații se realizează, deși este destul de dificil de realizat. În situația când se realizează analiza unei populații prin analiza fiecărui individ, știința ce descrie această populație se numește *Statistică Descriptivă*. Totuși, în cele mai multe cazuri, realizarea unei statistici

---

<sup>57</sup> De costuri, de control al rezultatelor etc.

descriptive nu se realizează. În aceste cazuri se folosesc unități statistice mai mici, pe baza lor putându-se realiza o extensie la populație. Acest gen de statistică se numește *Statistică Inferențială*. În secțiunea următoare vom prezenta diferite tipuri de unități statistice.

### 3.3. Eșantion. Lot

#### **Exemplul 42**

O firmă dorește să lanseze într-un anumit oraș un produs și ar dori să afle părerea populației din acel oraș, referitor la produs. Cel mai sigur ar fi să întrebe pe toți locuitorii din oraș, dar costurile, din punct de vedere financiar, ar fi destul de ridicate. Care ar fi soluția pentru a reduce costurile?

O definiție destul de uzuală este: eșantionul reprezintă subansamblul unei populații<sup>58</sup>. Pornind de la această definiție putem observa că un eșantion reprezintă mulțimea de elemente<sup>59</sup> care conțin una sau mai multe caracteristici comune. Desigur, ați observat că această definiție este asemănătoare cu cea a populației. Deci unde este diferența? Numărul de elemente, în cazul eșantionului, este mai mic decât cel al populației. Dar oare, având un grup elemente ce au aceeași caracteristică, putem să spunem ceva despre populație? Răspunsul este afirmativ doar dacă eșantionul este reprezentativ, dacă el conține la nivel micro toate elementele esențiale ce se regăsesc în toată populația.

#### **Exemplul 43**

Pornind de la ipoteza<sup>60</sup>: Cu cât crește vechimea în muncă cu atât dorința de schimbare scade. În acest caz, vom avea două eșantioane: Primul va fi format din muncitori angajați de curând, iar al doilea din muncitori ce au o vechime mai mare în muncă.

---

<sup>58</sup> Cf. [Rat].

<sup>59</sup> Reamintim că elementele pot fi persoane, obiecte, idei, opinii, evenimente etc.

<sup>60</sup> Întrucât nu constituie scopul acestei lucrări nu voi insista asupra problemelor legate de eșantionare.

Să presupunem că dorim să studiem dacă un fenomen poate influența o mulțime de elemente. Vom face măsurătorile pentru o mulțime înainte de petrecerea fenomenului și apoi, pe aceeași mulțime, vom face măsurători după petrecerea fenomenului. În acest caz nu o caracteristică realizează reunirea într-un grup a elementelor. Pentru acest gen de situații se folosește denumirea de *lot*. Primul lot<sup>61</sup> se va numi *lot de control*, iar cel de-al doilea lot<sup>62</sup> se va numi *lot experimental*.

#### **Exemplul 44**

Dorim să studiem dacă sistemul de notare cu note de la 1 la 100 ar fi mai bun decât sistemul de notare de la 1 la 10. Pentru aceasta vom lua două loturi. În primul lot vom avea elevii unei clase la care vom face notarea de la 1 la 10, iar al doilea lot vom avea elevii din altă clasă unde vom face notarea de la 1 la 100. În acest caz, am folosit denumirea de *lot* întrucât obiectul studiului îl reprezintă metoda de notare și nu elevii în sine<sup>63</sup>.

### **3.4. Eveniment**

#### **Exemplul 45**

Un operator trebuie să aplice un chestionar pe teren persoanelor peste 40 de ani de sex masculin. Având multe refuzuri, fiind destul de obosit, s-a dus acasă unde a rugat niște prieteni să-l ajute să „completeze” chestionarele. Cum s-a modificat evenimentul?

---

<sup>61</sup> Cel dinaintea petrecerii fenomenului.

<sup>62</sup> Cel de după ce fenomenul studiat a avut loc.

<sup>63</sup> Să observăm că obținerea de diferențe semnificative între cele 2 loturi nu înseamnă automat că un sistem de notare este mai bun decât celălalt.

### Exemplul 46

Un student avea ca temă studiul atenției concentrate la elevii de clasa a 3-a. În urma centralizării datelor a venit la profesorul îndrumător cu următorul tabel:

| Subiecți | Atenția concentrată | Vârsta |
|----------|---------------------|--------|
| B.B.     | 70                  | 9 ani  |
| T.E.     | 90                  | 9 ani  |
| C.A.     | 90                  | 9 ani  |
| R.I.     | 100                 | 10 ani |
| S.E.     | 90                  | 9 ani  |
| S.P.     | 90                  | 10 ani |
| P.V.     | 90                  | 9 ani  |
| L.I.     | 90                  | 9 ani  |
| B.D.     | 70                  | 10 ani |
| T.C.     | 90                  | 9 ani  |
| B.A.     | 90                  | 9 ani  |
| A.B.     | 100                 | 7 ani  |
| P.F.     | 90                  | 10 ani |
| B.D.     | 60                  | 9 ani  |
| R.C.     | 100                 | 9 ani  |
| I.V.     | 80                  | 9 ani  |
| C.A.     | 100                 | 8 ani  |
| S.S.     | 90                  | 13 ani |
| P.A.     | 100                 | 9 ani  |
| N.R.     | 90                  | 9 ani  |

În urma examinării acestui tabel, profesorul îi spune că trebuie să verifice datele și că tabelul nu e complet, deci proiectul nu poate fi notat.

Odată stabilit tipul de unitate statistică cu care vom lucra<sup>64</sup>, vom dori să realizăm diferite experimente, experiențe pe tipul de unitate statistică ales. În teoria probabilităților, rezultatul unei experiențe se numește *eveniment*.

---

<sup>64</sup> Populație, eșantion sau lot.

### **Exemplu 47**

Să presupunem că dorim să analizăm cât la sută din studenți vor intra în sala de curs în cele 10 minute dinaintea cursului. Populația va fi compusă din toți studenții, deoarece indiferent de an și de facultatea de unde sunt studenții, toți au ore de curs în programa de învățământ. Vom lua un lot<sup>65</sup> de la un curs dintr-o anumită zi. Ca eveniment (rezultat al experienței) putem avea 20 de persoane ce au intrat în ultimele 10 minute. Să notăm acest eveniment cu A. Cu B vom nota evenimentul 30 de persoane au intrat în ultimele 10 minute și cu C vom nota evenimentul 70 de persoane au intrat în ultimele 10 minute la curs.

Un alt tip de eveniment ar putea fi „în ultimele 10 minute să intre la curs mai puțin de 35 de persoane”. Pe acesta îl putem nota cu D. Se observă că evenimentul D include și evenimentul A și evenimentul B. Vom spune că evenimentele A și B sunt evenimente primare și evenimentul D este compus.

Ca tipuri de evenimente putem avea cele două cazuri extreme:

a) Evenimentul imposibil (sau evenimentul nul), care, așa cum se poate deduce din denumire, este acel eveniment care nu se poate produce. Acesta se poate obține prin îmbinarea a 2 evenimente care nu se pot produce în același timp.

### **Exemplul 48**

Să spunem că avem evenimentul A: să obținem mai mult de 7 puncte la un joc și evenimentul B: numărul obținut la aruncarea unui zar. Evenimentul rezultat din combinarea<sup>66</sup> celor două evenimente este C: la un joc să obținem mai mult de 7 puncte la aruncarea unui zar. Acest lucru evident este imposibil.

---

<sup>65</sup> În acest caz putem spune că este vorba de un lot deoarece nu ne interesează caracteristicile populației, ci fenomenul.

<sup>66</sup> Matematic vorbind este vorba de intersecția celor 2 evenimente.

b) Evenimentul total. Acest gen de eveniment este cel care se produce, în mod sigur, în urma unui eveniment.

În continuare vom prezenta principalii indicatori statistici ce se pot folosi în analiza primară a datelor.

Indicatorii tendinței centrale „sunt indicatori sintetici cu ajutorul cărora urmărim să exprimăm (...) ceea ce este tipic, esențial, stabil, obiectiv și caracteristic într-o serie de numere”<sup>67</sup>.

### 3.5. Minimul și maximul

La o primă privire asupra datelor de tip cantitativ putem deja observa doi indicatori: minimul și maximul.

#### Exemplul 49

Într-un liceu am studiat manifestarea complexului de inferioritate la un lot de 30 de adolescenți. Studiem posibilitatea ca să avem o manifestare a complexului de inferioritate prin valențe ridicate ale anxietății<sup>68</sup>.

Se observă în datele centralizate în tabelul de mai jos că maximul este 33, iar minimul este 19. Pornind de la aceste constatări am putea, de exemplu, să începem prin a investiga elevii ce au obținut aceste rezultate extreme.

|     | Nota A |     | Nota A |     | Nota A |
|-----|--------|-----|--------|-----|--------|
| S1  | 32     | S12 | 25     | S23 | 22     |
| S2  | 31     | S13 | 19     | S24 | 25     |
| S3  | 33     | S14 | 23     | S25 | 21     |
| S4  | 27     | S15 | 27     | S26 | 25     |
| S5  | 29     | S16 | 32     | S27 | 25     |
| S6  | 25     | S17 | 31     | S28 | 19     |
| S7  | 23     | S18 | 33     | S29 | 23     |
| S8  | 22     | S19 | 27     | S30 | 27     |
| S9  | 25     | S20 | 29     |     |        |
| S10 | 21     | S21 | 25     |     |        |
| S11 | 25     | S22 | 23     |     |        |

<sup>67</sup> Cf. [Țiț] p. 114.

<sup>68</sup> Am luat în considerare doar anxietatea manifestă.

### 3.6. Media

#### **Exemplul 50**

În urma aplicării unui test de memorie pe un lot de copii s-au obținut următoarele rezultate:

|    | <i>memoria vizuală</i> | <i>memoria auditivă</i> |
|----|------------------------|-------------------------|
| LM | 5                      | 6                       |
| VA | 3                      | 4                       |
| MA | 2                      | 4                       |
| TC | 4                      | 6                       |
| SA | 2                      | 2                       |
| TS | 2                      | 3                       |
| BR | 4                      | 5                       |
| FI | 4                      | 4                       |
| DC | 6                      | 8                       |
| CR | 1                      | 2                       |
| LL | 3                      | 4                       |

Care sunt valorile pentru care am putea spune că un copil are memorie vizuală bună în raport cu grupul studiat?

#### **Exemplul 51**

Într-o lucrare cu titlul „Legătura dintre mame și copiii de 6 ani”, studentul, la prezentarea lotului pe care a făcut acest studiu, a calculat media vârstelor componentilor lotului studiat. Este corect sau nu?

Media reprezintă unul dintre indicatorii cei mai folosiți pentru prezentarea datelor. Cea mai utilizată medie este media aritmetică. Din punct de vedere matematic ea se poate defini astfel:

Fiind dată o serie de  $n$  numere  $x_1, x_2, \dots, x_n$ , media aritmetică va fi:

$$m_x = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1)$$

Folosind notația  $x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$ , formula (1) se mai poate scrie:

$$m_x = \frac{\sum_{i=1}^n x_i}{n} \quad (2)$$

**Observație.** Dacă la aplicarea unui test obținem aceeași valoare pentru toți subiecții, să zicem 35, atunci media va chiar 35. Practic însă, un asemenea rezultat ar trebui să ne ducă la investigarea validității procedurii de aplicare al testului.

### Exemplul 52

În urma unei lucrări de control la o clasă elevii au obținut următoarele note: 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10. Pentru a realiza o evaluare a rezultatelor obținute profesorul dorește să calculeze media pe clasă la acel test. Pornind de la modul de abordare a datelor avem două posibilități de a calcula media la testul pe clasă.

Metoda 1. Fie datele prezentate liniar. 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 7, 7, 7, 7, 8, 8, 8, 9, 9, 9, 9, 10, 10, 10. Folosind formula (1) obținem  $m_x = 6,96$ .

Metoda 2. Putem pune date sub forma unui tabel:

|    |   |
|----|---|
| 4  | 3 |
| 5  | 5 |
| 6  | 2 |
| 7  | 4 |
| 8  | 3 |
| 9  | 4 |
| 10 | 3 |

Unde, pe prima coloană, am pus notele, iar pe coloana a doua am pus numărul de elevi corespunzători pentru fiecare notă. Dacă dorim să

calculăm media pentru a afla suma notelor putem calcula astfel  $4 \cdot 3 + 5 \cdot 5 + 6 \cdot 2 + 7 \cdot 4 + 8 \cdot 3 + 9 \cdot 4 + 10 \cdot 3$ , iar pentru a afla numărul de elevi putem aduna numerele din coloana a doua. Modul acesta de calcul este specific mediei ponderate. Căci media unui grup la un test este dat de notele și ponderile lor. Revenind la exemplul nostru luăm  $x_1 = 4, x_2 = 5, x_3 = 6, x_4 = 7, x_5 = 8, x_6 = 9, x_7 = 10$ , și ponderile  $p_1 = 3, p_2 = 5, p_3 = 2, p_4 = 4, p_5 = 3, p_6 = 4, p_7 = 3$ . Atunci folosind formula mediei ponderate este:

$$m_{px} = \frac{x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n p_n}{p_1 + p_2 + \dots + p_n} \quad (3)$$

vom obține  $m_{px} = 6,96$ , adică exact același rezultat ca la metoda 1.

### Exemplul 53

S-a aplicat un test de atenție concentrată la copiii din clasa a III-a. În urma testului s-au obținut următoarele rezultate:

| <i>Atenție concentrată-cantitate</i> |                              |
|--------------------------------------|------------------------------|
| <i>Valori</i>                        | <i>Ponderi</i> <sup>69</sup> |
| 60                                   | 1                            |
| 70                                   | 2                            |
| 80                                   | 1                            |
| 90                                   | 11                           |
| 100                                  | 7                            |

În acest caz media va fi:

$$m_{px} = (60 \cdot 1 + 70 \cdot 2 + 80 \cdot 1 + 90 \cdot 11 + 100 \cdot 7) / (1 + 2 + 1 + 11 + 7) = 1970 / 22 = 89,54$$

Deci se poate presupune, pentru această clasă, că pentru caracteristica *cantitate*, avem 4 elevi cu o atenție concentrată mică, 11 elevi au o atenție concentrată medie și 7 au atenție concentrată mare.

<sup>69</sup> Nr de elevi ce au obținut valoarea din stânga.

Problema, care se pune în această situație, este de a raporta media acestei clase la media populației alcătuită din elevii de clasa a III-a<sup>70</sup>.

În acest ultim exemplu am prezentat un alt rol al mediei: acela de indicator de poziție. Adică, raportându-ne la medie, putem spune dacă o valoare se află într-o anumită categorie sau în alta. De exemplu, calculând valoarea medie a înălțimii unui grup, putem defini ce înseamnă individ scund și ce înseamnă individ înalt pentru acel grup.

Mulți autori consideră că avantajul mediei constă în faptul că este ușor de calculat. Totuși, acest lucru poate constitui și un dezavantaj, deoarece, fiind ușor de calculat, comparativ cu alți indicatori, există tendința să se tragă concluzii în urma calculului mediei. Concluzii care pot fi eronate.

### 3.7. Cuantile

De multe ori, în practică, informații importante ne pot fi aduse de modul în care datele sunt distribuite între cea mai mare valoare și cea mai mică valoare. Deoarece poziția lor, în raport cu seria de numere, este bine determinată, acești indicatori se mai numesc și *indicatori de poziție*. De exemplu, putem scrie datele în ordine crescătoare și să vedem ce date se află la stânga mijlocului și ce date se află la dreapta mijlocului<sup>71</sup>. Valorile ce împart seria de numere în părți egale se numesc *cuantile*.

#### 3.7.1. Mediana

##### **Exemplul 54**

Studentii dintr-o grupă au vrut să-și aleagă un șef de grupă. Pentru că nu au ajuns la un consens, au rugat un profesor să-i ajute în această problemă. În urma discuțiilor avute cu studenții profesorul a determinat 3 candidați pentru această funcție. Apoi a pus pe fiecare student să acorde punctaje între 1 și 5 celor 3 candidați. S-au obținut următoarele date:

---

<sup>70</sup> Această medie a populației se poate găsi în scala de interpretare a testului.

<sup>71</sup> Așa cum vom vedea mai departe, acest „mijloc” se numește *mediană*.

|         | S10  | S15  | S22  |
|---------|------|------|------|
| S1      | 1    | 3    | 5    |
| S2      | 2    | 3    | 5    |
| S3      | 1    | 3    | 5    |
| S4      | 2    | 4    | 5    |
| S5      | 5    | 3    | 2    |
| S6      | 1    | 4    | 3    |
| S7      | 5    | 3    | 2    |
| S8      | 2    | 4    | 1    |
| S9      | 2    | 3    | 2    |
| S10     |      | 1    | 2    |
| S11     | 5    | 3    | 2    |
| S12     | 5    | 3    | 1    |
| S13     | 5    | 4    | 2    |
| S14     | 2    | 3    | 5    |
| S15     | 5    |      | 5    |
| S16     | 5    | 4    | 2    |
| S17     | 2    | 3    | 2    |
| S18     | 5    | 3    | 5    |
| S19     | 5    | 3    | 2    |
| S20     | 2    | 3    | 5    |
| S21     | 5    | 4    | 5    |
| S22     | 2    | 3    |      |
|         |      |      |      |
| media   | 3.29 | 3.19 | 3.24 |
| mediana | 2    | 3    | 2    |

Cine credeți că este mai indicat să fie ales șef al grupei?

*Mediana* este cea cuantilă ce împarte o serie de numere în două părți egale. Vom nota mediana cu  $M_e$ . Modul de determinare al mediane se poate realiza în două moduri (în funcție de numărul de elemente):

1) Dacă avem un număr impar de valori, atunci mediana va fi cea care are numărul de elemente de la stânga egal cu numărul de elemente de la dreapta ei.

### Exemplul 55

Într-o echipă A, s-a cerut muncitorilor să noteze șeful cu note de la 1 la 10. În urma centralizării datelor s-au obținut următoarele rezultate:

|               | AM1 | AM2 | AM3 | AM4 | AM5 |
|---------------|-----|-----|-----|-----|-----|
| Note acordate | 1   | 2   | 2   | 3   | 10  |

În acest caz, mediana este cea din dreptul lui AM3 și  $M_e = 2$ . Eticheta AM3 se mai numește, în acest caz, *element median*<sup>72</sup>.

2) Dacă avem un număr par de elemente, atunci mediana este egală cu media celor două elemente de la mijlocul seriei de numere ordonate.

### Exemplul 56

Analog cu problema pusă la exemplul 55 s-a realizat o cercetare și la o echipă B:

|               | BM1 | BM2 | BM3 | BM4 | BM5 | BM6 |
|---------------|-----|-----|-----|-----|-----|-----|
| Note acordate | 1   | 1   | 2   | 3   | 3   | 3   |

Mediana este între BM3 și BM4, deci pentru a o calcula vom face media între valorile corespunzătoare acestor etichete. Deci  $M_e = 2,5$

Care este rolul medianei? Mediana este utilă când, în analiza procesului, intervin păreri subiective și există pericolul ca valorile extreme să influențeze rezultatul<sup>73</sup>. Reluând exemplele 54 și 55, să presupunem că dorim să studiem următoarea problemă: în care echipă șeful de echipă este mai bine văzut de către subalterni? Dacă am folosi media observăm că șeful echipei A obține media 3, în timp ce șeful echipei B obține media 2,2. Deci am putea concluziona că șeful echipei A

---

<sup>72</sup> În [Rot] este numit *indivîd median*, p. 46.

<sup>73</sup> Se pot folosi mai multe metode în momentul când avem de analizat punctaje rezultate în urma unor păreri subiective. De exemplu, la notele acordate la concursurile de patinaj se elimină cea mai mare și cea mai mică valoare și se realizează media valorilor rămase.

este mai bine văzut de subalterni. Folosind mediana, șeful echipei A obține valoarea 2, în timp ce șeful echipei B obține valoarea 3.

Asimetria unei serii de numere o putem determina prin studierea poziției medianei față de medie. În capitolul anterior, în finalul secțiunii 2.3.1., am văzut cum se reprezintă cele două extreme (date grupate la cea mai mică valoare, respectiv la cea mai mare valoare). În practică, aceste cazuri le putem întâlni foarte rar.

Ne-ar putea interesa însă situațiile în care seria de numere tinde spre o extremă sau alta. Când avem curba situată mai spre stânga datelor, vom spune că avem o asimetrie spre stânga, ceea ce în practică ne va indica tendința valorilor de a se grupa spre cele mai mici valori. Analog se poate concluziona atunci când avem curba datelor spre dreapta.

**Exemplul 57**

Într-un studiu comparativ s-a aplicat un test de evaluare a singurătății emoționale<sup>74</sup> pe două loturi de tineri: primul lot a fost format din tineri cu studii superioare<sup>75</sup>, iar al doilea din tineri ce au doar studii liceale<sup>76</sup>. În urma centralizării datelor s-au obținut următoarele rezultate:

| TSL | TSS | TSL | TSS | TSL | TSS |
|-----|-----|-----|-----|-----|-----|
| 8   | 6   | 0   | 10  | 7   | 3   |
| 4   | 5   | 14  | 3   | 21  | 9   |
| 6   | 12  | 1   | 11  | 10  | 11  |
| 3   | 10  | 13  | 10  | 27  | 9   |
| 3   | 12  | 16  | 12  | 11  | 20  |
| 19  | 11  | 6   | 13  | 13  | 6   |
| 7   | 10  | 16  | 13  | 7   | 6   |
| 5   | 9   | 6   | 10  | 2   | 6   |
| 12  | 6   | 5   | 3   | 1   | 3   |
| 4   | 10  | 9   | 3   | 1   | 4   |

<sup>74</sup> Esli b.

<sup>75</sup> Notăți în continuare cu TSS.

<sup>76</sup> Notăți în continuare cu TSL.

În urma calculelor, pentru TSL s-a obținut media  $m_{TSL} = 8,57$  și mediana  $M_{eTSL} = 7,00$  și pentru TSS s-a obținut media  $m_{TSS} = 8,53$  și mediana  $M_{eTSS} = 9,50$ . În acest caz se observă că, deși, din punct de vedere al mediei, cele două loturi sunt sensibil apropiate, mediana ne aduce o nouă informație: cei din grupul TSL au tendință spre valori mici, pe când cei din TSS au o tendință spre valorile mari.

### 3.7.2. Cuartilele

Cuartilele<sup>77</sup> sunt acei indicatori ce împart datele în patru părți egale. Se poate observa ușor că cea de-a doua cuartilă reprezintă, de fapt, mediana. Pentru a obține cele patru cuartile folosim următoarea metodă. Calculăm întâi mediana și rezultatul îl vom nota cu Q2. Apoi, pentru datele din stânga lui Q2 calculăm mediana și obținem valoarea Q1. Analog pentru datele din dreapta lui Q2 calculăm din nou mediana și obținem Q3. Aparent, această metodă pare complicată, dar deoarece există destul de multe softuri ce au o funcție ce poate calcula această mediană<sup>78</sup>, lucrurile devin destul de simple.

#### Exemplu 58

Mai jos am calculat cuartilele pentru șirul de date. Am obținut mediana egală cu 7,00. Ordonăm crescător datele și pentru șirul astfel obținut calculăm Q1 și Q3:

|     |   |   |   |   |   |    |   |   |   |   |   |    |
|-----|---|---|---|---|---|----|---|---|---|---|---|----|
| TSL | 0 | 1 | 1 | 2 | 3 | 4  | 5 | 6 | 6 | 6 | 7 |    |
|     |   |   |   |   |   | Q1 |   |   |   |   |   | Q2 |

|   |   |    |    |    |    |    |    |    |    |
|---|---|----|----|----|----|----|----|----|----|
| 7 | 8 | 10 | 11 | 12 | 13 | 13 | 16 | 16 | 19 |
|   |   |    |    | Q3 |    |    |    |    |    |

**Observație.** Am calculat mediana pentru un număr par de date și atunci mediana va fi egală cu media valorilor a 15-a și a 16-a. În

<sup>77</sup> Acest noțiuni trebuie tratate cu atenție întrucât din punct de vedere al denumirii sunt foarte asemănătoare. Reamintim faptul că prin *cuantile* înțelegem acele puncte din serie ce împart seria de numere în părți egale.

<sup>78</sup> De exemplu, în Excel avem funcția MEDIAN.

acest caz cele două valori sunt egale și valoarea lor este 7, deci și mediana va fi egală cu 7. Pentru calculul mediane primelor 11 date<sup>79</sup> mediana va fi numărul „de la mijloc”, adică a șaptea valoare. Deci, Q1 va fi 4. Analog se va calcula Q3 care va fi egal cu 12.

### 3.7.3. Valoare modală<sup>80</sup>

Se numește *valoare modală* a unei serii de numere acea valoare ce apare de mai multe ori în acea serie. Vom nota valoarea modală cu  $vm$ .

**Observație.** În general, când se folosește termenul de *valoare* ne gândim la valoarea exprimată prin numere. Așa cum la obiecte există în afară de valoarea din punct de vedere financiar și alt gen de valori (de exemplu, *valoarea emoțională*), și în această secțiune vom folosi sensul mai larg al termenului de *valoare* (vezi exemplul 43)

#### Exemplul 43 (continuare)

Se observă că pentru datele

| Atenție concentrată-cantitate |                       |
|-------------------------------|-----------------------|
| Valori                        | Ponderi <sup>81</sup> |
| 60                            | 1                     |
| 70                            | 2                     |
| 80                            | 1                     |
| 90                            | 11                    |
| 100                           | 7                     |

avem  $vm = 90$ .

**Observație.** Se observă că pentru acest caz avem media egală cu 89,54, iar mediana este  $M_e = 90 = vm$ .

<sup>79</sup> Adică Q1.

<sup>80</sup> Am preferat această denumire celorlalte denumiri cum ar fi *modul* sau *moda*, pentru a evita confuziile.

<sup>81</sup> Nr de elevi ce au obținut valoarea din stânga.

### 3.8. Indicatori ai dispersiei

#### **Exemplul 59**

Un student a realizat un studiu pe elevii dintr-o școală. Pentru a putea trage o concluzie asupra comportamentului copiilor în lotul studiat a luat elevii din clasele a 2-a, a 5-a și a 8-a. De ce îndrumătorul i-a spus să-și modifice modul de alcătuire al lotului?

În analiza unui grup ne interesează tendința grupului, privit ca un element. Dacă, de exemplu, avem un lot de subiecți care, în medie, a obținut la un test un anumit punctaj, atunci putem trage o concluzie asupra comportamentului grupului.

#### **Exemplul 60**

În luna mai s-a aplicat unui grup de elevi din clasa a 8-a un test de anxietate. Punctajul la acest test este cuprins între 0 și 10 unde, cu cât punctajul este mai mare, cu atât anxietatea este mai puternică. În urma prelucrării datelor media lotului de elevi a fost 6,7. Analizând media, rezultă că tendința generală este că grupul este, în mod esențial, normal din punct de vedere al anxietății, cu o ușoară tendință spre categoria „puțin nevrotic”.

În momentul în care însă analizăm un grup prin media lui, trebuie să ținem cont de faptul că media nu este singura caracteristică a unui lot de subiecți și că putem avea aceeași medie la două loturi și totuși, analizând acele loturi, să observăm diferențe importante între ele. Uneori, în practică, deși avem aceeași medie pentru două eșantioane, ele sunt diferite din punct de vedere al tendinței<sup>82</sup>.

Pentru a exemplifica, fie următoarea situație:

Măsurăm înălțimea populației cu vârste între un an și 60 de ani în București. În acest caz obținem o medie a acestui lot de subiecți, însă această medie nu ar avea nicio relevanță din cauza variabilității mari a valorilor.

---

<sup>82</sup> Reamintim că tendința era dată de poziția medianei față de medie.

### 3.8.1. Amplitudinea

Amplitudinea este egală cu diferența dintre cea mai mare și cea mai mică valoare. Matematic, aceasta se scrie altfel:

$$A = x_{\max} - x_{\min} \quad (4)$$

#### Exemplul 49 (continuare)

|     | Nota A |     | Nota A |     | Nota A |
|-----|--------|-----|--------|-----|--------|
| S1  | 32     | S12 | 25     | S23 | 22     |
| S2  | 31     | S13 | 19     | S24 | 25     |
| S3  | 33     | S14 | 23     | S25 | 21     |
| S4  | 27     | S15 | 27     | S26 | 25     |
| S5  | 29     | S16 | 32     | S27 | 25     |
| S6  | 25     | S17 | 31     | S28 | 19     |
| S7  | 23     | S18 | 33     | S29 | 23     |
| S8  | 22     | S19 | 27     | S30 | 27     |
| S9  | 25     | S20 | 29     |     |        |
| S10 | 21     | S21 | 25     |     |        |
| S11 | 25     | S22 | 23     |     |        |

Avem  $x_{\max} = 33$  și  $x_{\min} = 19$ . Deci amplitudinea va fi  

$$A = 33 - 19 = 14.$$

În practică, uneori prin calculul raportului  $\frac{x_{\max}}{x_{\min}}$  se pot obține mai multe informații. Acest raport se poate folosi în studiul proceselor dinamice, pentru a afla de câte ori este mai mare valoarea maximă față de valoarea minimă.

### 3.8.2. Abaterea medie ( $A_m(a)$ )

Ne poate interesa distanța dintre o valoare calculată și o valoare dată. În termeni statistici rezultatul acestei scăderi se va numi *abaterea valorii calculate de la valoarea dată*.

#### Exemplul 61

Se aplică un test unui pacient înainte și după aplicarea unui tratament. Înainte de tratament a obținut 35 de puncte, iar după aplicarea tratamentului a obținut

valoarea 29. Știind că valoarea critică este 40, calculăm abaterea față de această valoare pentru a realiza o primă apreciere asupra fenomenului. Abaterea<sup>83</sup> dintre punctajul obținut înainte de tratament față de valoarea critică este 5, în timp ce abaterea dintre punctajul obținut după tratament și valoarea critică este 11, deci există posibilitatea ca tratamentul să aibă un efect pozitiv.

**Observație.** Am spus că există posibilitatea ca tratamentul să aibă un efect pozitiv în general, deoarece în urma aplicării unui experiment unui singur subiect, nu putem decât să emitem ipoteze.

Să presupunem că dorim să studiem mai în amănunt problema pusă în exemplul 3.9. Pentru aceasta ar trebui să realizăm un studiu pe mai mulți subiecți. În acest caz, notând cu  $x_1, x_2, \dots, x_n$  valorile obținute de pacienți înainte de tratament și cu  $y_1, y_2, \dots, y_n$  valorile obținute de pacienți după tratament. Pentru a compara abaterile față de valoarea critică o soluție simplă este să calculăm o medie a abaterilor înainte de tratament și după tratament. Această medie se numește **abaterea medie** și se notează  $A_m(a)$ <sup>84</sup>

### Exemplul 62

În urma aplicării testului pe mai mulți subiecți înainte și după tratament s-au obținut următoarele rezultate:

| Nr. subiect | Înainte de tratament | Abaterea față de 40 | După tratament | Abaterea față de 40 |
|-------------|----------------------|---------------------|----------------|---------------------|
| 1           | 35                   | 5                   | 29             | 11                  |
| 2           | 36                   | 4                   | 32             | 8                   |
| 3           | 33                   | 7                   | 32             | 8                   |
| 4           | 30                   | 10                  | 29             | 11                  |
| 5           | 36                   | 4                   | 32             | 8                   |
| 6           | 35                   | 5                   | 33             | 7                   |
| 7           | 29                   | 11                  | 29             | 11                  |
| 8           | 32                   | 8                   | 30             | 10                  |
| $A_m(40) =$ |                      | <b>6.75</b>         | $A_m(40) =$    | <b>9.25</b>         |

În practică, în cele mai multe cazuri, se calculează media abaterilor față de medie.

<sup>83</sup> Val crit-val exp = 40 – 35.

<sup>84</sup> Abaterea medie față de valoarea  $a$ .

### Exemplul 63

S-a studiat variația anxietății, seara și dimineața, pe un lot de 10 studenți aflați în sesiunea de examene. O primă parte a acestui experiment a fost măsurarea anxietății voalate a studenților dimineața. În urma centralizării datelor s-au obținut rezultatele:

|                 | <i>anxietate<br/>voalată</i> |                 | <i>abaterea față de<br/>medie (<math>x - m</math>)</i> |
|-----------------|------------------------------|-----------------|--------------------------------------------------------|
| x1              | 12                           |                 | 4                                                      |
| x2              | 14                           |                 | 2                                                      |
| x3              | 20                           |                 | 4                                                      |
| x4              | 20                           |                 | 4                                                      |
| x5              | 12                           |                 | 4                                                      |
| x6              | 16                           |                 | 0                                                      |
| x7              | 18                           |                 | 2                                                      |
| x8              | 17                           |                 | 1                                                      |
| x9              | 15                           |                 | 1                                                      |
| x10             | 16                           |                 | 0                                                      |
| <b>medie(m)</b> | <b>16.00</b>                 | <b>abaterea</b> | <b>2.20</b>                                            |

### Observație

Având în vedere modul de definire al abaterii, am luat distanța dintre valori și medie în modul.

#### 3.8.3. Varianța

Un alt indicator al dispersiei ce se folosește în practică este varianța. Aceasta se exprimă astfel:

$$s^2 = \frac{(x_1 - m)^2 + (x_2 - m)^2 + \dots + (x_n - m)^2}{n - 1} \quad (6)$$

unde  $m$  reprezintă media valorilor  $x_1, x_2, \dots, x_n$ . Această formulă este valabilă pentru eșantion.

**Observație:** Termenul de *varianță* a fost introdus pentru prima oară de Ronald Fisher în 1918 în lucrarea *The Correlation Between Relatives on the Supposition of Mendelian Inheritance*.

### 3.8.4. Coeficient de Omogenitate (variabilitate)

În cazul când alegem un grup, singura problemă este să ne asigurăm că acel grup se manifestă unitar din punctul de vedere al caracteristicii studiate. Pentru a studia aceasta ne vom raporta la omogenitatea datelor care se calculează cu ajutorul coeficientului de variabilitate. Formula este:

$$CV = \frac{100s}{m_x} \% \quad (7)$$

Avem următoarea interpretare:

- Dacă  $CV < 10\%$  atunci datele sunt foarte omogene
- Dacă  $10\% < CV < 20\%$  datele sunt omogene
- Dacă  $CV > 20\%$  atunci datele nu sunt omogene, deci nu se comportă unitar în raport cu caracteristica studiată și, în acest caz, fie refacem testarea pentru alt grup, fie măsurăm o altă caracteristică ce ar putea caracteriza grupul.

#### **Exemplul 64**

Pe un lot de studenți în sesiune s-a măsurat anxietatea voalată dimineața și seara. În urma centralizării datelor s-a obținut următorul tabel.

| <i>dimineața</i> | <i>seara</i> |
|------------------|--------------|
| 21               | 20           |
| 16               | 17           |
| 15               | 16           |
| 24               | 22           |
| 16               | 18           |
| 12               | 18           |
| 23               | 23           |
| 15               | 16           |
| 16               | 17           |
| 18               | 19           |
| 24               | 26           |
| 21               | 20           |
| 13               | 15           |

|         |          |         |
|---------|----------|---------|
|         | 23       | 22      |
|         | 16       | 17      |
|         | 14       | 16      |
|         | 14       | 15      |
|         | 12       | 14      |
|         | 19       | 18      |
|         | 15       | 16      |
| medie   | 17.35    | 18.25   |
| abt std | 4.003617 | 3.09286 |
| CV      | 23.1%    | 16.9%   |

Analizând rezultatele obținute la coeficientul de variabilitate, observăm că datele obținute dimineața pentru anxietatea voalată nu sunt omogene ( $CV > 20\%$ ). Deci subiecții, din acest punct de vedere sunt destul de variați. Însă seara datele devin omogene ( $CV < 20\%$ ), ceea ce arată că, din punct de vedere al anxietății, subiecții ajung să reacționeze la fel. Deci am putea concluziona că, în sesiune, seara, anxietatea voalată caracterizează grupul studiat.

În cele prezentate în acest capitol referirile erau făcute relativ la un grup de subiecți. Problema care se pune, în cazul acestor analize, este să determinăm dacă concluziile obținute asupra acestui grup se transferă și asupra populației din care face parte grupul. Concluziile la care am putea ajunge, privitor la anxietatea grupului de studenți, ar mai fi valabile dacă am avea mai mulți studenți? Dar dacă ne interesează să vedem gradul de anxietate al studenților și să comparăm rezultatele obținute dimineața cu cele obținute seara? Concret, în acest caz, avem două grupuri de studenți – un prim grup format din reacția studenților dimineața și un al doilea grup format din reacția studenților seara. Pentru a compara aceste două grupuri, indicatorii prezentați până acum nu sunt suficienți. Ceea ce ne va ajuta, în acest gen de analize, sunt testele statistice.

**Observații.** În cele ce urmează se va înțelege prin analiza a două grupuri următoarele:

1. un grup să fie populația, iar al doilea grup să fie un eșantion. În această situație, ceea ce ne interesează este dacă eșantionul face parte din acea populație.
2. un grup să fie un eșantion, iar al doilea grup un alt eșantion. În această situație, ceea ce ne interesează este dacă cele două eșantioane fac parte din aceeași populație.

Faptul că un eșantion face parte dintr-o populație sau că două eșantioane sunt din aceeași populație reprezintă ipoteza de plecare în folosirea testului. Generalizând, putem spune că avem o ipoteză nulă, de obicei aceasta se notează cu  $H_0$ , în care se presupune că între cele două grupuri nu avem diferențe semnificative. Atunci când  $H_0$  nu este îndeplinită spunem că s-a realizat ipoteza alternativă, notată cu  $H_a$  (eșantionul nu face parte din populație sau cele două eșantioane nu fac parte din aceeași populație).

Bineînțeles atunci când se emite o ipoteză trebuie să știm la ce ne așteptăm dacă ea se verifică. Deși pare un lucru evident am ținut totuși să scriu această regulă deoarece adeseori studenții, în emiterea ipotezelor, nu țin cont de această regulă. Pentru a înțelege mai bine aceasta dăm exemplul de mai jos:

### **Exemplul 65**

La susținerea unui proiect un student și alege următoarea temă: *Influența vârstei asupra anxietății pre și postoperatorii*. Și-a ales cele două loturi și a observat că această diferență se manifestă la bărbați înainte de operație în funcție de vârstă. Când i s-a pus întrebarea ce înseamnă această diferență, studentul nu a știut ce să răspundă.

Un alt concept ce ne interesează atunci când dorim să utilizăm un test este *nivelul de semnificație*. Acesta ne arată care este probabilitatea de la care evenimentul (fenomenul) poate fi considerat imposibil. În general apar următoarele valori:

- $\alpha = 0,05$  (sau 5%)
- $\alpha = 0,01$  (sau 1%)
- $\alpha = 0,001$  (sau 0,1%)

Altfel spus, de exemplu  $\alpha = 0,05$  semnifică faptul că maxim 5 cazuri din 100 pot să nu respecte  $H_0$ , pentru a considera totuși că  $H_0$  este valabilă pentru lotul de 100 de subiecți.

Nivelul de semnificație este stabilit de utilizator în funcție de precizia pe care o dorește pentru fenomenul studiat.

### 3.9. Compararea mediilor

#### **Exemplul 66**

Un student își propune să studieze dacă din punct de vedere al atenției involuntare există diferențe între fetele și băieții dintr-o clasă a patra. Ce ar trebui să facă pentru aceasta?

#### **Exemplul 67**

Se poate aplica un test de comparare a mediilor la prelucrarea unui chestionar?

Unul dintre matematicienii care s-a ocupat de problema comparării statistice a mediilor este William Sealy Gosset, cunoscut și sub numele de „Student”. Patronul fabricii de bere Guinness din Dublin, Claude Guinness avea ca politică a firmei sale angajarea a celor mai buni absolvenți de la Universitățile din Oxford și Cambridge pentru a aplica în procesele industriale de la Guinness noțiuni de biochimie și statistică. W.Gosset a publicat prima lucrare despre testul  $t$  în revista *Biometrika* în 1908 sub pseudonimul „Student” deoarece în contractul semnat cu fabrica de bere era stipulat că metodele statistice sunt secret de serviciu. De aceea, identitatea lui Student nu a fost dezvăluită multă vreme pentru a nu fi acuzat de încălcarea clauzelor contractuale.

#### 3.9.1. Compararea mediei unui grup cu o valoare dată

În cele ce urmează, prin *grade de libertate* se va înțelege valoarea dată de numărul de subiecți minus doi. De exemplu, dacă aplicăm un test de comparare a mediilor între două loturi, un lot de 15 subiecți și un al doilea lot de 19 subiecți atunci gradele de libertate vor avea date de  $15+19-2 = 32$ .

Atunci când se aplică un test se obține o valoare calculată care se compară cu o valoare tabelară. Valoarea tabelară depinde de nivelul de semnificație pe care îl alegem și de gradele de libertate.

Uneori, în practică, avem nevoie să comparăm media obținută prin calcul cu o valoare dată. Concret, ceea ce se poate lua ca ipoteză în această situație este dacă între media grupului și media unei populații, ce are ca medie valoarea dată, există diferențe.

### **Exemplul 68**

Într-o școală s-a dat un test la 30 de elevi. Media clasei a fost 7,30. Știind că la nivelul capitalei media obținută a fost 7,20 se pune problema dacă media clasei studiate este semnificativ mai mare decât media notelor obținute la nivelul capitalei.

Notăm cu  $m$  media grupului, cu  $n$  numărul de subiecți din grup, cu  $v$  valoarea dată și cu  $s$  abaterea medie pătratică a grupului. Atunci valoarea calculată  $z$  va fi dată de formula:

$$z = \frac{m - v}{\frac{s}{\sqrt{n}}} \quad (8)$$

Această valoare se va compara cu valoarea tabelară și dacă  $z < z_{tab}$  atunci ipoteza nulă este acceptată, adică se poate presupune ca  $m = v$  (media grupului este egală cu media populației).

### **3.9.2. Testarea ipotezei privind diferența dintre mediile a două grupuri cu număr redus de subiecți**

Pentru a testa ipoteza privind diferența dintre mediile a două grupuri de volum redus pornim de la presupunerile că avem următoarele condiții îndeplinite:

- cele două grupuri provin din populații normal distribuite
- dispersiile corespunzătoare celor două grupuri sunt egale
- selecția elementelor unui grup s-a realizat independent de selecția celui alt grup.

În aceste condiții, enumerate mai sus, se realizează, mai întâi, o aproximare a dispersiei conform următoarelor formule:

$$s_c^2 = \frac{(n_1 - 1)^2 s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2} \tag{9}$$

În această situație, vom avea ipoteza  $H_0$ : diferența între medii este zero. Pentru a verifica această ipoteză vom calcula:

$$t_{calc} = \frac{|m_1 - m_2|}{\sqrt{s_c^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \tag{10}$$

Analog cu procedeul de la secțiunea 3.9.1., și aici avem o valoare tabelară,  $t_{tabel}$ , care se găsește în tabele statistice și depinde de nivelul de semnificație ales de utilizator și de gradele de libertate. Dacă  $t_{calc} < t_{tabel}$  atunci este acceptată ipoteza, adică diferențele dintre cele două medii sunt nesemnificative.

**Exemplul 69**

În urma aplicării testului Guilford de abilități ale gândirii divergente (flexibilitatea) la un liceu s-au obținut următoarele rezultate:

| Băieți |    | Fete |    |    |
|--------|----|------|----|----|
| 5      | 7  | 8    | 6  | 14 |
| 11     | 5  | 5    | 5  | 7  |
| 11     | 11 | 12   | 8  | 7  |
| 5      | 7  | 7    | 6  | 9  |
| 7      | 7  | 11   | 3  | 9  |
| 6      | 7  | 5    | 6  |    |
| 13     | 8  | 7    | 8  |    |
| 6      | 12 | 5    | 6  |    |
| 8      | 12 | 8    | 7  |    |
| 16     | 10 | 10   | 16 |    |
| 7      | 11 | 14   | 12 |    |
| 6      | 7  | 7    | 6  |    |

Calculând mediile, obținem la băieți valoarea 8,54, iar la fete valoarea 8,07. Problema care ne-o punem este dacă diferența de 0,47 este semnificativă.

Calculăm valoarea  $t_{\text{calc}}$  și obținem  $t_{\text{calc}} = 0,117$ . Din tabel obținem pentru un nivel de semnificație de 0,05 o valoare  $t_{\text{tabel}} = 2,007$ . Cum  $t_{\text{tabel}} > t_{\text{calc}}$  înseamnă că ipoteza nulă este acceptată, adică nu avem diferențe semnificative între fete și băieți din punct de vedere al flexibilității. Adică flexibilitatea la vârsta de 17 ani în liceul studiat nu depinde de sexul subiecților.

### 3.10. Testul $\chi^2$

Exemple unde se poate folosi acest gen de test:

#### **Exemplul 70**

În anul 2000, cercetătorii de la *National Opinion Research Center* (NORC), de la Universitatea din Chicago au realizat un sondaj pe un eșantion de 1768 de persoane (994 femei și 774 bărbați). Una din întrebări s-a referit la faptul dacă respondenții sunt de acord cu avortul sau nu (indiferent de motiv). În urma centralizării rezultatelor s-au obținut următoarele rezultate:

|                  | <i>Masculin</i> | <i>Feminin</i> |
|------------------|-----------------|----------------|
| De acord         | 309             | 396            |
| Nu sunt de acord | 465             | 598            |

Putem spune că diferențele între cele două loturi sunt ne semnificative sau putem afirma că femeile sunt într-o mai mare măsură împotriva avortului (indiferent de motiv)?

### Exemplul 71<sup>85</sup>

La 31 martie 2003 la nivelul țării 13,60% din persoanele cu handicap erau copii în timp ce persoane adulte cu handicap erau 86,4%. În aceeași perioadă în Municipiul București erau înregistrați 3427 de copii cu handicap și 29.833 de adulți. Diferă situația din București de cea la nivelul întregii țări?

Să presupunem că dorim să facem observații asupra unor categorii. De exemplu, într-un liceu dorim să observăm modul în care elevii de clasa a 9-a se acomodează cu cerințele liceului. Să presupunem că avem: prima categorie – cei care s-au acomodat, a doua – cei care au un nivel mediu de acomodare și a treia categorie – cei care nu s-au acomodat. În urma unor observații din anii anteriori avem următoarele frecvențe: la prima categorie avem  $m_1$  elevi, la a doua categorie avem  $m_2$  elevi, la a treia categorie avem  $m_3$  elevi. Problema pe care ne-o punem este dacă într-un anumit liceu avem cam aceeași distribuție a frecvențelor. Dacă vom nota cu  $x_1$ ,  $x_2$  și  $x_3$  diferențele apărute față de ceea ce se știa din anii anteriori atunci vom spune că la prima categorie am obținut  $m_1 + x_1$ , la a doua categorie am obținut  $m_2 + x_2$  și la a treia categorie am obținut  $m_3 + x_3$ . Vom spune atunci că diferențele între cele două distribuții vor fi mai mici cu cât valorile  $x_i$  sunt mai mici. R.A. Fisher, în lucrarea sa *Statistical methods for research workers*, notează cu  $\chi^2$  următorul număr

$$\chi^2 = \text{Sum} \left( \frac{x^2}{m} \right) = \frac{x_1^2}{m_1} + \frac{x_2^2}{m_2} + \frac{x_3^2}{m_3} + \dots \quad (11)$$

Se observă că există o interdependență între diminuarea lui  $x$  și diminuarea lui  $\chi^2$  (cu cât  $x$  este mai mic cu atât  $\chi^2$  este mai mic). Deci, putem afirma: cu cât  $\chi^2$  este mai mic cu atât distribuțiile vor fi mai apropiate. Analog cu testul de comparație a mediilor și aici avem o ipoteză nulă:

$H_0$  : distribuțiile frecvențelor nu diferă

și o ipoteză alternativă:

$H_a$ : distribuțiile frecvențelor diferă semnificativ

---

<sup>85</sup> Sursa: *Anuarul Statistic* ediția martie 2003.

Aceste ipoteze le vom verifica folosindu-ne de valoarea lui  $\chi^2$ . În cele de mai sus am prezentat intuitiv modul de calcul al valorii  $\chi^2_{calc}$ , valoare pe care o vom compara cu o valoare tabelară,  $\chi^2_{tabel}$ , care depinde de nivelul de semnificație ales și gradele de libertate. În acest caz, gradele de libertate sunt date de: (numărul de categorii – 1) x (numărul de loturi observate – 1). Deci: dacă  $\chi^2_{calc} < \chi^2_{tabel}$  spunem că se acceptă ipoteza nulă (nu avem diferențe semnificative între loturile comparate), iar dacă  $\chi^2_{calc} > \chi^2_{tabel}$  atunci ipoteza nulă este respinsă și se acceptă ipoteza alternativă (avem diferențe semnificative între loturile comparate).

### **Exemplul 72**

La un sondaj realizat în 2004 pe un eșantion de 1500, printre altele, s-au adresat următoarele întrebări cu variantele de răspuns:

1. Câtă încredere aveți în Biserică?

- a. Foarte multă
- b. Multă
- c. Puțină
- d. Foarte puțină
- e. Deloc

2. Câtă încredere aveți în armată?

- a. Foarte multă
- b. Multă
- c. Puțină
- d. Foarte puțină
- e. Deloc

În urma centralizării rezultatelor s-a obținut următorul tabel:

|               | <i>F.<br/>multă</i> | <i>Multă</i> | <i>Puțină</i> | <i>F.<br/>puțină</i> | <i>Deloc</i> | <i>total</i> |
|---------------|---------------------|--------------|---------------|----------------------|--------------|--------------|
| Bise-<br>rica | 626                 | 525          | 201           | 58                   | 90           | 1500         |
| Arma-<br>ta   | 377                 | 694          | 289           | 54                   | 86           | 1500         |

Calculând obținem:  $\chi^2_{calc} = 232,89$ . Luând ca nivel de semnificație 5% și gradele de libertate  $(5-1) \cdot (2-1) = 4$ , obținem  $\chi^2_{tabel} = 9,49$ . Deoarece  $\chi^2_{calc} > \chi^2_{tabel}$  se respinge ipoteza nulă și se acceptă ipoteza alternativă, adică avem diferențe semnificative între cele două distribuții.

### **Observații:**

- La aplicarea testului  $\chi^2$  vom lucra cu frecvențe absolute, nu frecvențe relative (nu procente).
- Dacă dorim să comparăm distribuțiile a două eșantioane putem considera distribuția unuia ca fiind distribuție teoretică, iar a celuilalt distribuție experimentală.

## **3.11. Coeficientul de corelație**

### **Exemplul 73**

La aplicarea testului pentru identificarea trăsăturilor accentuate pe un lot de 18 de subiecți dependenți de heroină s-au obținut următoarele rezultate:

|          | <b>Dem<br/>I</b> | <b>Ex<br/>VIII</b> |
|----------|------------------|--------------------|
| C. M.    | 83.00            |                    |
| Ș. P.    | 16.60            | 75.00              |
| G. C.    | 33.20            | 75.00              |
| Ghe. C   | 24.90            | 50.00              |
| E. E     | 8.30             |                    |
| P. A.    | 41.50            | 75.00              |
| S. I.    | 49.80            | 50.00              |
| R. S.    | 74.70            | 100.00             |
| C. D.    | 16.00            | 50.00              |
| I. M.    | 24.90            | 75.00              |
| V. F. B. | 8.30             | 25.00              |
| C. B.    | 91.30            | 75.00              |
| M. D.    | 33.20            | 100.00             |
| S. C. B. | 8.30             |                    |
| Ș.B.     | 16.60            | 75.00              |
| C.A.L.   | 33.20            | 25.00              |
| S.D.     | 33.20            | 25.00              |
| B.M.C.   | 24.90            | 50.00              |

Așa cum se poate observa din tabel pe coloana în care s-a măsurat exaltarea sunt niște date lipsă. Există vreo posibilitate de a putea estima aceste date?

### Exemplul 74

Un student a venit cu tabelul de mai jos și m-a rugat să-l ajut să interpreteze valoarea coeficientului de corelație pentru cele două șiruri a cărui valoare calculată este de  $-0,38$ .

| <i>Atenția voluntară III A</i> | <i>Atenția voluntară III B</i> |
|--------------------------------|--------------------------------|
| 7                              | 10                             |
| 10                             | 10                             |
| 10                             | 5                              |
| 10                             | 9                              |
| 10                             | 2                              |
| 5                              | 10                             |
| 10                             | 4                              |
| 10                             | 9                              |
| 10                             | 5                              |
| 10                             | 5                              |
| 10                             | 10                             |
| 10                             | 5                              |
| 10                             | 10                             |
| 10                             | 10                             |
| 10                             | 10                             |
| 9                              | 10                             |
| 9                              | 10                             |
| 10                             | 10                             |
| 10                             | 10                             |
| 10                             | 10                             |
| 5                              | 10                             |

De ce lucrul pe care-l cere acest student nu se poate realiza?

### Exemplul 75

La un proiect un student are următoarele ipoteze de lucru :

- cu cât oamenii sunt mai extrovertiți, cu atât asertivitatea lor este mai mare.
- cu cât personalitatea de TIP A predomină față de cea de TIP B, cu atât nivelul stresului este mai mare.
- dacă simptomele psihologice ale stresului sunt mai însemnate, atunci nivelul stresului este crescut.

Poate folosi corelația pentru a studia aceste trei ipoteze ?

Atunci când pe un lot de subiecți studiem mai multe variabile se pune problema evaluării relației dintre aceste variabile, chiar dacă nu presupunem că ar putea exista o dependență între aceste variabile. Problema corelației a fost analizată de Galton în urma unui studiu realizat asupra eredității caracterelor, studiu apărut la mijlocul secolului al XIX-lea.

O altă lucrare importantă în cercetarea corelației o reprezintă studiul realizat de Pearson și Lee, studiu în care au calculat corelații dintre statură, mărimea palmei și lungimea antebrațului. Acest studiu a fost realizat pe un grup de 1401 de familii engleze.

### Exemplu 76 (Pearson și Lee – modificat)

S-a măsurat pe un lot de 10 familii statura fraților și surorilor în inci. După transformarea în cm se obține următorul tabel:

| <i>Nr familiei</i> | <i>Sora</i> | <i>Frate</i> |
|--------------------|-------------|--------------|
| 1                  | 175         | 180          |
| 2                  | 163         | 173          |
| 3                  | 165         | 168          |
| 4                  | 160         | 170          |
| 5                  | 165         | 178          |
| 6                  | 157         | 180          |
| 7                  | 165         | 178          |
| 8                  | 163         | 185          |
| 9                  | 168         | 183          |
| 10                 | 150         | 165          |

În urma calculelor s-a obținut un coeficient de corelație  $r = 0,56$

Ca și la teste, și în cazul coeficientului de corelație se pune problema de a determina semnificabilitatea. Pentru aceasta luăm ca ipoteză  $H_0$  : variabilele sunt independente, iar ca ipoteză alternativă  $H_a$ : variabilele sunt dependente. Pentru a verifica  $H_0$  trebuie să vedem dacă coeficientul de corelație este 0. Pentru aceasta se calculează valoarea  $t_{calc}$ :

$$t_{calc} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} \quad (12)$$

Această valoare se compară cu valoarea  $t_{tab}$  de la testul Student pentru  $n-2$  grade de libertate. La fel ca la comparația mediilor vom avea:

- dacă  $t_{calc} < t_{tab}$  atunci diferențele sunt nesemnificative, deci putem spune că nu avem corelație semnificativă între cele două variabile
- dacă  $t_{calc} > t_{tab}$  atunci diferențele sunt semnificative, deci putem spune că este semnificativă corelația obținută.

**Observație.** Având în vedere că pentru calculul semnificabilității corelației  $t_{tab}$  depinde de numărul de subiecți, putem observa că o valoare a corelației poate fi semnificativă pentru un număr de subiecți, dar nesemnificativă pentru un număr mai mic de subiecți.

Dacă în cazul testelor prezentate anterior finalitatea este destul de clară, în cazul coeficientului de corelație lucrurile nu mai sunt la fel de clare. La ce folosește acest coeficient de corelație? Pentru a ne apropia de răspunsul la această întrebare amintim cele spuse de R.A. Fisher despre coeficientul de corelație: „în munca experimentală locul lui este mai puțin central; s-a dovedit a fi util în stadiul explorator al unei cercetări, ca de exemplu când doi factori care se presupunea a fi independenți par a fi asociați în apariția lor; dar rareori, în condiții experimentale controlate se dorește ca să exprimăm concluziile sub forma unui coeficient de corelație”<sup>86</sup>. Am dat acest citat pentru a arăta grija cu care recomandă Fisher folosirea acestui coeficient. Tocmai pentru că este relativ ușor de folosit (mai ales atunci când folosim un computer pentru prelucrarea datelor) și foarte puternic în a da explicații unor rezultate la care nu se vede rezultatul imediat, recomandăm și noi

---

<sup>86</sup> În lucrarea *Statistical methods for research workers*, cap. VI.

folosirea lui doar într-o fază incipientă a cercetării și pentru eventuale validări a unor concluzii la care s-a ajuns folosind și alte instrumente.

Una din greșelile frecvente ce apar în interpretarea coeficientului de corelație este ca atunci când s-au găsit două variabile ce se corelează să se pună imediat o relație între ele de tip cauză efect. Pentru a înțelege cele expuse mai sus vom da un exemplu:

### ***Exemplul 77***

În perioada 1928-1934 s-a realizat un studiu în care s-au măsurat în fiecare an numărul celor internați la spitalele psihiatrice și numărul celor care au cumpărat aparate de radio. S-a observat că între cele două variabile exista o corelație foarte bună de 0,94. În mod evident însă nu este vorba de o relație de cauză efect între cele două fenomene, ba chiar mai mult, nici măcar nu se poate afirma că între cele două variabile există o legătură reală. Normal, la vremea respectivă aparatul de radio era un lucru nou și bineînțeles că numărul de cumpărători să crească, cât despre numărul crescut al celor care au fost internați în spitalele de psihiatrie mai degrabă situația politico-economică a acelor vremuri să fie cauza reală a creșterii numărului de pacienți ai spitalelor de psihiatrie.

Deci am văzut cum nu trebuie folosit coeficientul de corelație. Dar întrebarea pusă inițial devine: „De ce e nevoie de acest coeficient care nu ajută la găsirea unor motivații și care, în plus, poate duce și la erori privind dependența aparentă a două variabile?” Pentru a răspunde la această întrebare ne vom referi la regresia liniară.

### **3.12. Regresia liniară**

Dacă într-o populație avem două variabile despre care știm că au o corelație semnificativă, atunci:

- când creșterea uneia implică creșterea celeilalte avem corelație pozitivă
- când creșterea unei variabile implică descreșterea celeilalte variabile avem corelație negativă.

În oricare din aceste cazuri semnificația acestui fenomen statistic este că putem exprima o variabilă în funcție de cealaltă, adică faptul că ar exista o funcție care poate exprima valoarea unei variabile în funcție de cealaltă variabilă. Din această cauză a apărut confuzia legată de ideea de relație de tip cauză-efect între variabilele corelate. În fapt, este vorba că există o variabilă ușor observabilă și în funcție de aceasta putem calcula o altă variabilă mai greu observabilă. Funcția aceasta se numește *regresie*. Dacă funcția este de forma unei funcții de gradul întâi, atunci vom spune că regresia este liniară.

### **Exemplul 78**

Pe un grup de 20 de elevi s-a studiat corelația dintre memoria auditivă și mediile obținute pe semestrul 1. În urma centralizării datelor s-au obținut următoarele rezultate:

|     | <i>Memoria auditivă</i> | <i>Note</i> |
|-----|-------------------------|-------------|
| S1  | 30                      | 6,25        |
| S2  | 25                      | 9,10        |
| S3  | 10                      | 5,42        |
| S4  | 20                      | 7,11        |
| S5  | 70                      | 9,15        |
| S6  | 20                      | 5,10        |
| S7  | 50                      | 8,16        |
| S8  | 50                      | 6,41        |
| S9  | -                       | 5,30        |
| S10 | 20                      | 5,32        |
| S11 | 0                       | 6,33        |
| S12 | 50                      | 9,15        |
| S13 | -                       | 7,55        |
| S14 | 30                      | 8,16        |
| S15 | 50                      | 9,85        |
| S16 | 75                      | 8,65        |
| S17 | -                       | 4,65        |
| S18 | 0                       | 6,35        |
| S19 | 10                      | 5,45        |
| S20 | -                       | 7,56        |
| S21 | 10                      | 5,12        |
| S22 | 10                      | 8,14        |
| S23 | -                       | 8,15        |
| S24 | 20                      | 4,20        |
| S25 | 10                      | 6,48        |

Se observă că 5 elevi nu au răspuns la testul pentru memoria auditivă. Întrucât coeficientul de corelație obținut pentru 20 de elevi este 0,638 și este semnificativ, putem calcula o funcție de regresie cu ajutorul căreia să putem estima scorurile la memoria auditivă pe baza mediilor și pentru elevii care au lipsit atunci când a fost aplicat acest test.

În exemplul de mai sus, variabila mai ușor observabilă este variabila *note*<sup>87</sup> și, de aceea, o putem numi *variabilă independentă*, în timp ce variabila cu scorurile obținute de elevi la testul de *memorie auditivă*, necesită aplicarea unui test și o putem mult mai ușor obține prin estimarea cu ajutorul funcției de regresie. De aceea, putem numi variabila scoruri test *memorie auditivă*, variabilă dependentă. Însă locul acestor variabile se poate schimba. Știind că există o corelație semnificativă între notele obținute de elevi și scorurile la testul de *memorie auditivă*, putem aplica acest test la începutul unui semestru și putem estima pe baza lui care ar fi elevii predispuși să aibă rezultate școlare slabe. În acest al doilea caz, variabila scoruri obținute la testul de *memorie auditivă* devine variabilă independentă și variabila *note* devine variabila dependentă.

Cu ajutorul acestui exemplu am dorit de asemenea să arătăm și de ce nu este întotdeauna corect să se analizeze o corelație ca un indicator al unui fenomen de tip cauză-efect.

Un alt rol al corelației este și de a putea elimina anumite variabile. Să presupunem că într-un fenomen studiat intervin mai multe variabile. În statistică este știut că un număr mai mare de măsurători înseamnă, în primul rând, pericolul de a avea erori mari de calcul<sup>88</sup>. Dacă două variabile A și B au o corelație semnificativă, atunci când analizăm fenomenul respectiv, putem elimina una din ele. De obicei, se elimină variabila care se măsoară mai greu. Să presupunem că am eliminat variabila B atunci influența acestei

---

<sup>87</sup>Deoarece aflăm rezultatele din consultarea mediilor din catalogul de note, deci interacțiunea cu elevii este practic zero.

<sup>88</sup>Pe de altă parte un număr mai mare de măsurători efectuate înseamnă și cheltuieli mai mari (atât din punct de vedere financiar, cât și ca timp).

variabile în fenomen nu este eliminată, ci apare indirect prin intermediul variabilei A.

### **3.13. Rezolvarea exercițiilor de la începutul fiecărei secțiuni din capitolul 3**

În capitolul 3 la începutul fiecărei secțiuni am pus niște probleme cu trimitere la ceea ce urmează să se discute în acea secțiune. În continuare vom prezenta răspunsurile la problemele puse. Desigur, în cadrul discuțiilor de la seminar se pot găsi și alte soluții la problemele date.

#### **3.13.1. Caracteristici (variabile)**

##### ***Exemplul 34 (rezolvare)***

Pentru proiectul de diplomă un student și-a ales tema: *Complexul de inferioritate la adolescenți*. Ce ar trebui să cerceteze studentul în cadrul acestui proiect?

*Deoarece complexul de inferioritate se manifestă pregnant prin valențe ridicate ale anxietății și depresiei, cercetarea ar putea porni de la aceste două variabile*

##### ***Exemplul 35 (rezolvare)***

La două clase a V-a din aceeași școală, existau performanțe școlare semnificativ diferite, una din clase având rezultate foarte bune, în timp ce la cealaltă clasă, deși aveau cam aceeași profesori, problema principală a elevilor era să nu rămână corigenți. Pe ce ar trebui să se axeze un eventual studiu de determinare a cauzelor diferențelor între performanțele școlare ale elevilor?

*Desigur, putem avea multe puncte de pornire. O soluție ar fi să se ia variabile diferite tipuri de atenție: atenția concentrată, atenția vizuală etc.*

### 3.13.2. Populație

#### **Exemplul 39 (rezolvare)**

Pentru realizarea unui sondaj privitor la alegerile locale dintr-un anumit oraș, din cine se compune mulțimea pe care o studiem? Dar dacă sondajul s-ar referi la alegerile generale?

*În primul caz, populația va fi alcătuită din locuitorii din acel oraș cu drept de vot (practic, însă se vor lua locuitorii care au peste 18 ani). În cazul alegerilor locale populația se va compune din cetățenii cu drept de vot din toată țara. Ca observație s-ar putea include și cetățenii care au 17 ani, dacă alegerile ar avea loc peste un an.*

### 3.13.3. Eșantion. Lot

#### **Exemplul 42 (rezolvare)**

O firmă dorește să lanseze într-un anumit oraș un produs și ar dori să afle părerea populației din acel oraș referitor la produs. Cel mai sigur ar fi să întrebe pe toți locuitorii din oraș, dar costurile, și ca timp și financiare, ar fi destul de ridicate. Care ar fi soluția pentru a reduce costurile?

*În acest caz, răspunsul este evident: este vorba de a realiza un eșantion reprezentativ.*

### 3.13.4. Eveniment

#### **Exemplul 45 (rezolvare)**

Un operator trebuie să aplice un chestionar pe teren persoanelor peste 40 de ani de sex masculin. Având multe refuzuri, fiind destul de obosit, s-a dus acasă unde a rugat niște prieteni să-l ajute „completeze” chestionarele. Cum s-a modificat evenimentul?

*Evenimentul inițial se baza pe faptul că cei chestionați aveau vârsta de peste 40 de ani. În condițiile în care cei care completează nu îndeplinesc această condiție, deja avem un alt eveniment ale cărui rezultate nu vor reflecta datele evenimentului cerut. În plus, mai apare influența indirectă a operatorului, dată de faptul că cei care completeau făceau parte dintr-o mulțime restrânsă, cea a celor care erau prieteni cu operatorul.*

#### **Exemplul 46 (rezolvare)**

Un student avea ca temă studiul atenției concentrate la elevii de clasa a 3-a. În urma centralizării datelor a venit la profesorul îndrumător cu următorul tabel:

| <i>Subiecți</i> | <i>Atenția concentrată</i> | <i>Vârsta</i> |
|-----------------|----------------------------|---------------|
| B.B.            | 70                         | 9 ani         |
| T.E.            | 90                         | 9 ani         |
| C.A.            | 90                         | 9 ani         |
| R.I.            | 100                        | 10 ani        |
| S.E.            | 90                         | 9 ani         |
| S.P.            | 90                         | 10 ani        |
| P.V.            | 90                         | 9 ani         |
| L.I.            | 90                         | 9 ani         |
| B.D.            | 70                         | 10 ani        |
| T.C.            | 90                         | 9 ani         |
| B.A.            | 90                         | 9 ani         |
| A.B.            | 100                        | 7 ani         |
| P.F.            | 90                         | 10 ani        |
| B.D.            | 60                         | 9 ani         |
| R.C.            | 100                        | 9 ani         |
| I.V.            | 80                         | 9 ani         |
| C.A.            | 100                        | 8 ani         |
| S.S.            | 90                         | 13 ani        |
| P.A.            | 100                        | 9 ani         |
| N.R.            | 90                         | 9 ani         |

În urma examinării acestui tabel, profesorul îi spune că trebuie să verifice datele și că tabelul nu e complet, deci proiectul nu poate fi notat.

*O primă problemă este dată de faptul că atenția concentrată are două componente: calitativă și cantitativă, și din tabel nu reiese la care din acestea două s-a referit studentul când a realizat tabelul. Apoi, în lotul cu care lucrează apar cel puțin două cazuri ciudate: un elev de 7 ani și unul de 13 ani. Dar s-ar putea ca această ultimă eroare să fie doar o eroare de tastare. Un factor care nu influențează major experimentul este faptul că studentul nu are în coloana a treia numere, ceea ce îi va îngreuna calculele în cazul în care ar dori să studieze dacă există vreo legătură între vârsta elevilor și scorurile obținute la atenție.*

### 3.13.5. Media

#### **Exemplul 50 (rezolvare)**

În urma aplicării unui test de memorie pe un lot de copii s-au obținut următoarele rezultate:

|    | <i>Memoria vizuală</i> | <i>Memoria auditivă</i> |
|----|------------------------|-------------------------|
| LM | 5                      | 6                       |
| VA | 3                      | 4                       |
| MA | 2                      | 5                       |
| TC | 4                      | 6                       |
| SA | 2                      | 2                       |
| TS | 2                      | 3                       |
| BR | 4                      | 5                       |
| FI | 5                      | 3                       |
| DC | 6                      | 8                       |
| CR | 1                      | 2                       |
| LL | 3                      | 4                       |

Care sunt valorile pentru care am putea spune că un copil are memorie vizuală bună, în raport cu grupul studiat? Dar memorie auditivă?

*Cea mai simplă metodă de a afla care copii au memorie vizuală bună este de a calcula media pentru acest lot. În urma calculelor obținem valoarea mediei de 3,27 pentru memoria vizuală și 4,36 pentru memoria auditivă. De exemplu, despre subiectul FI putem spune că are memorie vizuală bună, dar memorie auditivă slabă, în timp ce despre subiectul MA putem spune că are memorie vizuală slabă, dar memorie auditivă bună. Acest gen de analiză este util atunci când dorim să monitorizăm un subiect în raport cu grupul din care face parte.*

#### **Exemplul 51 (rezolvare)**

Într-o lucrare cu titlul *Legătura dintre mame și copii de 6 ani*, studentul, la prezentarea lotului pe care a făcut-o în acest studiu, a calculat media vârstelor componentilor lotului studiat. Este corect sau nu?

*Aceasta este una din greșelile clasice pe care am putea-o denumi „să facem ceva statistic la cercetare”. Media vârstelor mamelor ar putea avea o relevanță, media copiilor ar trebui să fie foarte aproape de 6 ani, în schimb media obținută, punând la un loc vârstele copiilor și ale mamelor, să zicem că ar fi 18, nu arată absolut nimic, ci reprezintă doar un număr între vârsta copiilor și vârsta mamelor, număr fără nicio relevanță.*

### 3.13.6. Mediana

#### **Exemplul 54 (rezolvare)**

Studentii dintr-o grupă au vrut să-și aleagă un șef de grupă. Pentru că nu au ajuns la un consens, au rugat un profesor să-i ajute în această problemă. În urma discuțiilor avute cu studenții profesorul a determinat 3 candidați pentru aceasta funcție. Apoi a pus pe fiecare student să acorde punctaje între 1 și 5 celor 3 candidați. S-au obținut următoarele date:

|         | S10  | S15  | S22  |
|---------|------|------|------|
| S1      | 1    | 3    | 5    |
| S2      | 2    | 3    | 5    |
| S3      | 1    | 3    | 5    |
| S4      | 2    | 4    | 5    |
| S5      | 5    | 3    | 2    |
| S6      | 1    | 4    | 3    |
| S7      | 5    | 3    | 2    |
| S8      | 2    | 4    | 1    |
| S9      | 2    | 3    | 2    |
| S10     |      | 1    | 2    |
| S11     | 5    | 3    | 2    |
| S12     | 5    | 3    | 1    |
| S13     | 5    | 4    | 2    |
| S14     | 2    | 3    | 5    |
| S15     | 5    |      | 5    |
| S16     | 5    | 4    | 2    |
| S17     | 2    | 3    | 2    |
| S18     | 5    | 3    | 5    |
| S19     | 5    | 3    | 2    |
| S20     | 2    | 3    | 5    |
| S21     | 5    | 4    | 5    |
| S22     | 2    | 3    |      |
|         |      |      |      |
| media   | 3.29 | 3.19 | 3.24 |
| mediana | 2    | 3    | 2    |

Cine credeți că este mai indicat să fie ales șef al grupei?

*Se observă că între medii diferențele sunt foarte mici. Desigur s-ar putea alege ca șef de grupă s10 pentru că ar avea media cea mai mare <sup>89</sup>. Mediana, fiind indicatorul părerii subiective a grupului, constituie un reper mult mai bun. Deci, având în vedere că s15 are pentru mediană valoarea cea mai mare, profesorul o să recomande ca șef de grupă pe s15.*

### 3.13.7. Indicatori ai dispersiei

#### **Exemplul 59 (rezolvare)**

Un student a realizat un studiu pe elevii dintr-o școală. Pentru a putea trage o concluzie asupra comportamentului copiilor în lotul studiat a luat elevii din clasele a 2-a, a 5-a și a 8-a. De ce îndrumătorul i-a spus să-și modifice modul de alcătuire al lotului?

*În fapt este vorba de o problemă ce apare din cauza modului incorect de folosire al variabilei. În general, o variabilă este studiată pe un anumit lot dacă există posibilitatea ca ea să caracterizeze, într-un fel, acel lot. Deci, odată culese datele, se va obține o împrăștiere foarte mare a datelor. În acest caz se poate realiza un studiu comparativ între elevii celor trei ani de studiu.*

---

<sup>89</sup> Așa cum însă se observă din secțiunea cu testele de comparare a mediilor diferența între medii este nesemnificativă.

### 3.13.8. Compararea mediilor

#### **Exemplul 66 (rezolvare)**

Un student își propune să studieze dacă din punct de vedere al atenției involuntare există diferențe între fetele și băieții dintr-o clasă a patra. Ce ar trebui să facă pentru aceasta?

*Pentru a studia acest lucru studentul trebuie să alcătuiască două loturi de subiecți: unul cu fetele de clasa a patra și altul cu băieții de clasa a patra. Pentru a elimina orice alți factori de influență, trebuie să ia eleve și elevi din aceeași clasă (ca să aibă aceeași învățătoare).*

#### **Exemplul 67 (rezolvare)**

Se poate aplica un test de comparare a mediilor la prelucrarea unui chestionar?

*În general, nu se poate aplica un test de comparare a mediilor la un chestionar. Dar există posibilitatea unei întrebări în care se cere subiecților să dea note pentru a-și exprima opțiunea. În acest caz, întrebarea se interpretează prin calculul mediei. Deci, se poate folosi, în acest caz, testul de comparare a mediilor.*

### 3.13.9. Testul $\chi^2$

#### **Exemplul 70 (rezolvare)**

În anul 2000, cercetătorii de la *National Opinion Research Center* (NORC), de la Universitatea din Chicago au realizat un sondaj pe un eșantion de 1768 de persoane (994 femei și 774 bărbați). Una din întrebări s-a referit la faptul dacă respondenții sunt de acord cu avortul sau nu (indiferent de motiv). În urma centralizării rezultatelor s-au obținut următoarele rezultate:

|                  | Masculin | Feminin |
|------------------|----------|---------|
| De acord         | 309      | 396     |
| Nu sunt de acord | 465      | 598     |

Putem spune că diferențele între cele două loturi sunt nesemnificative sau putem afirma ca femeile sunt într-o mai mare măsură împotriva avortului (indiferent de motiv)?

*Dacă facem o comparație directă a datelor am putea crede că femeile sunt într-o măsură mai mare împotriva avortului. Dar dacă exprimăm datele în procente observăm că diferențele sunt nesemnificative:*

|                  | Masculin | Feminin |
|------------------|----------|---------|
| De acord         | 39.9%    | 39.8%   |
| Nu sunt de acord | 60.1%    | 60.2%   |

*O metodă mai sigură de a vedea acestea este folosirea testului  $\chi^2$ .*

### Exemplu 71 (rezolvare)

La 31 martie 2003 la nivelul țării 13,60% din persoanele cu handicap erau copii în timp ce persoane adulte cu handicap erau 86,4%. În aceeași perioadă în Municipiul București erau înregistrați 3427 de copii cu handicap și 29.833 de adulți. Diferă situația din București de cea la nivelul întregii țări?<sup>90</sup>

*Pentru a compara distribuția de la nivelul Municipiului București cu distribuția pe țară vom transforma în frecvențe absolute toate datele și vom obține următorul tabel:*

|                   | Copii | Adulți | Total |
|-------------------|-------|--------|-------|
| în România        | 3427  | 29833  | 33260 |
| în Mun. București | 4523  | 28737  | 33260 |

<sup>90</sup>Sursa: Anuarul Statistic ediția martie 2003.

<sup>90</sup>Sursa: Anuarul Statistic ediția martie 2003

Aplicând testul  $\chi^2$ , se vede că avem diferențe semnificative între distribuții. Deci am putea spune că la nivelul țării avem un procent mai mare de copii cu handicap decât în București, raportat la numărul total de persoane cu handicap.

### 3.13.10. Coeficientul de corelație

#### **Exemplul 73 (rezolvare)**

La aplicarea testului pentru identificarea trăsăturilor accentuate pe un lot de 18 de subiecți dependenți de heroină s-au obținut următoarele rezultate:

|          | <b>Dem</b> | <b>Ex</b>   |
|----------|------------|-------------|
|          | <b>I</b>   | <b>VIII</b> |
| C. M.    | 83.00      |             |
| Ș. P.    | 16.60      | 75.00       |
| G. C.    | 33.20      | 75.00       |
| Ghe. C   | 24.90      | 50.00       |
| E. E     | 8.30       |             |
| P. A.    | 41.50      | 75.00       |
| S. I.    | 49.80      | 50.00       |
| R. S.    | 74.70      | 100.00      |
| C. D.    | 16.00      | 50.00       |
| I. M.    | 24.90      | 75.00       |
| V. F. B. | 8.30       | 25.00       |
| C. B.    | 91.30      | 75.00       |
| M. D.    | 33.20      | 100.00      |
| S. C. B. | 8.30       |             |
| Ș.B.     | 16.60      | 75.00       |
| C.A.L.   | 33.20      | 25.00       |
| S.D.     | 33.20      | 25.00       |
| B.M.C.   | 24.90      | 50.00       |

Așa cum se poate observa din tabel pe coloana în care s-a măsurat exaltarea sunt niște date lipsă. Există vreo posibilitate de a putea estima aceste date?

*Acest exemplu este reluat la partea în care se prezintă regresia liniară. Calculând coeficientul de corelație pentru datele complete, se observă că avem o corelație semnificativă, ceea ce înseamnă că putem realiza o estimare a datelor lipsă cu ajutorul regresiei liniare.*

#### **Exemplul 74**

Un student a venit cu tabelul de mai jos și m-a rugat să-l ajut să interpreteze valoarea coeficientului de corelație pentru cele două șiruri a cărui valoare calculată este de -0,38.

| Atenția voluntară III A | Atenția voluntară III B |
|-------------------------|-------------------------|
| 7                       | 10                      |
| 10                      | 10                      |
| 10                      | 5                       |
| 10                      | 9                       |
| 10                      | 2                       |
| 5                       | 10                      |
| 10                      | 4                       |
| 10                      | 9                       |
| 10                      | 5                       |
| 10                      | 5                       |
| 10                      | 10                      |
| 10                      | 5                       |
| 10                      | 10                      |
| 10                      | 10                      |
| 9                       | 10                      |
| 9                       | 10                      |
| 10                      | 10                      |
| 10                      | 10                      |
| 10                      | 10                      |
| 5                       | 10                      |

De ce lucrul pe care-l cere acest student nu se poate realiza?

*Acesta este un exemplu de folosire defectuoasă a coeficientului de corelație. Prin definiție acest coeficient ne arată legătura dintre două caracteristici ale aceluiași grup studiat. Ori, în acest caz, este vorba de aceeași caracteristică, aplicată pe două grupuri diferite. Ceea ce se poate aplica este comparația mediilor celor două grupuri.*

### **Exemplul 76**

La un proiect un student are următoarele ipoteze de lucru :

- cu cât oamenii sunt mai extrovertiți, cu atât asertivitatea lor este mai mare.
- cu cât personalitatea de TIP A predomină față de cea de TIP B, cu atât nivelul stresului este mai mare.
- dacă simptomele psihologice ale stresului sunt mai însemnate, atunci nivelul stresului este crescut.

Poate folosi corelația pentru a studia aceste trei ipoteze ?

*Ipotezele ce le putem verifica cu ajutorul coeficientului de corelație sunt cele de forma „cu cât ..., cu atât”. Deci primele două ipoteze se verifică cu ajutorul coeficientului de corelație. Cea de-a treia ipoteză nu este bine formulată deoarece enunțul este confuz („mai însemnate”).*

## 4. NOȚIUNI INTRODUCTIVE DESPRE GRAFURI ȘI BAZE DE DATE

### 4.1. Grafuri

Noțiunea de *categorie* este legată de cea de *graf* și „că, odată cu o colecție de obiecte”, [trebuie] „să fie considerate modalitățile de interconectare a obiectelor respective”<sup>91</sup>. De aceea, grafurile pot găsi o ilustrare în domeniul social, unde analizarea unui grup nu reprezintă doar suma analizelor indivizilor din grup, ci și interacțiunea dintre membrii grupului. Este un lucru îndeobște recunoscut că același grup poate reacționa într-un fel când are un lider și altfel când are alt lider. Sau, evident, într-un mod acționează cineva ca lider al unui grup și altfel ca simplu membru al grupului.

În acest capitol ne propunem doar să deschidem o ușă către posibilitatea de a folosi grafurile atât ca reprezentare a unor fenomene, cât și ca un posibil instrument în modelarea unor fenomene sociale. De aceea, ne vom limita la o prezentare intuitivă a unor noțiuni legate de teoria grafurilor.

#### 4.1.1. Scurt istoric

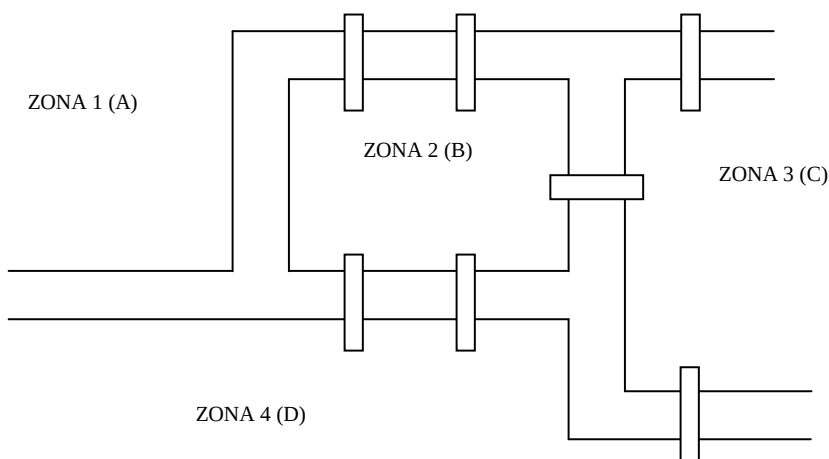
În secolul al XVIII-lea orașul Königsberg (Kalingrad astăzi) era traversat de râul Pregel care împărțea orașul în patru zone, acestea fiind unite prin 7 poduri. Schematic, harta orașului se poate reprezenta astfel:

---

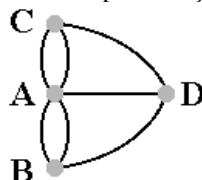
<sup>91</sup> Cf. [Stăn], pg

---

<sup>91</sup> Cf. [Stăn].



Locuitorii orașului și-au pus problema dacă, în plimbările pe care le făceau duminică prin oraș, este posibil ca pornind dintr-o parte a orașului să poată traversa toate podurile orașului fără a trece de două ori pe același pod și să se întoarcă la punctul de plecare. Problema a fost prima oară rezolvată de matematicianul Leonard Euler. Figura de mai sus se poate pune și sub forma unui graf. Pentru aceasta vom nota zona 1 cu A, zona 2 cu B, zona 3 cu C și zona 4 cu D. Reprezentarea va fi:



Relațiile dintre oameni le putem reprezenta cu ajutorul grafurilor.

#### 4.1.1. Definiția unui graf

Un graf este o structură compusă din noduri<sup>92</sup> și arce<sup>93</sup>. Grafurile se pot împărți în următoarele două categorii:

- grafuri neorientate
- grafuri orientate

---

<sup>92</sup> În unele manuale se folosește denumirea de vârfuri.

<sup>93</sup> Sau muchii

Diferența dintre cele două tipuri de grafice, constând în faptul că, în timp ce la grafurile neorientate aveam doar noduri unite prin

<sup>92</sup> În unele manuale se folosește denumirea de *vârfuri*.

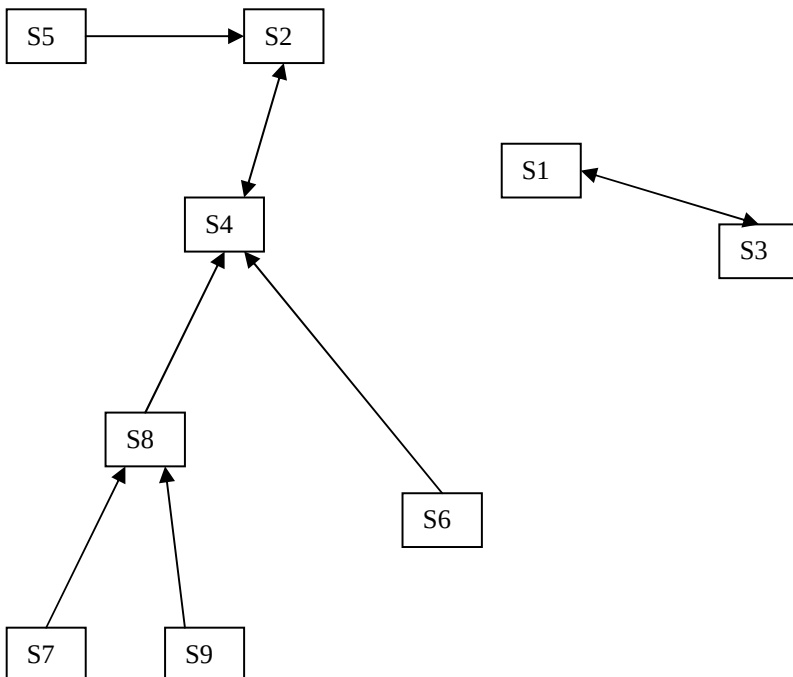
<sup>93</sup> Sau *muchii*.

arce, fără a avea un sens de parcurgere între noduri, la grafurile orientate avem în plus și un sens de parcurgere între noduri.

În exemplul prezentat în secțiunea 4.1. avem un graf neorientat, unde A, B, C și D sunt noduri, iar liniile care unesc aceste noduri sunt arcele grafului.

### **Exemplul 32 (continuare)**

Reluăm exemplul 32 din capitolul 2. Considerând subiecții ca fiind noduri și interacțiunile dintre ei fiind arcele grafului, putem reprezenta relațiile dintre ei sub forma unui graf astfel:

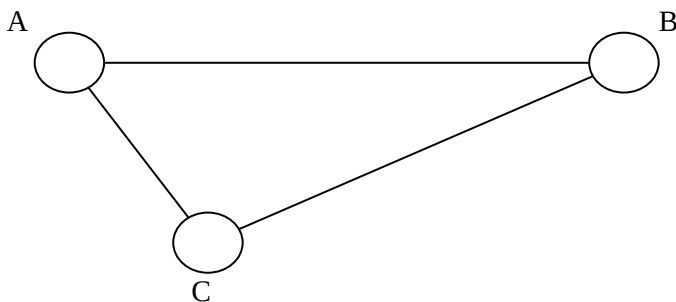


Se observă că avem, în fapt, două grafuri și o primă concluzie ar fi că subiecții S1 și S3 nu interacționează în niciun mod cu restul grupului. Acest mod de reprezentare poate releva și alte aspecte ale relațiilor dintr-un grup decât reprezentarea clasică a unei sociograme. Tot la această problemă putem avea și un alt gen de reprezentare în care facem o reprezentare prin grafuri pentru interacțiunile pozitive între membrii grupului și o altă reprezentare pentru interacțiunile negative în grupul studiat.

#### 4.1.3. Clasificarea grafurilor

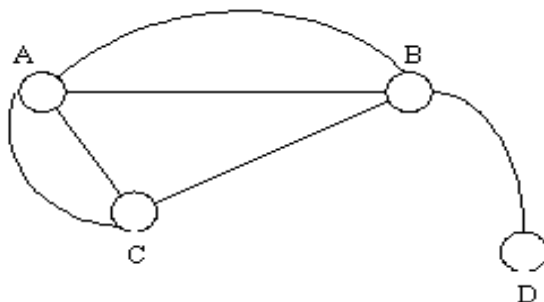
O altă clasificare a grafurilor este dată de forma lor. Astfel, putem avea următoarele tipuri de grafuri:

Grafuri simple – grafuri în care între orice două noduri avem un unic arc:



*Între A și C, A și B și B și C avem câte un singur arc*

Grafuri multiple – grafuri în care între două noduri putem avea mai multe arce



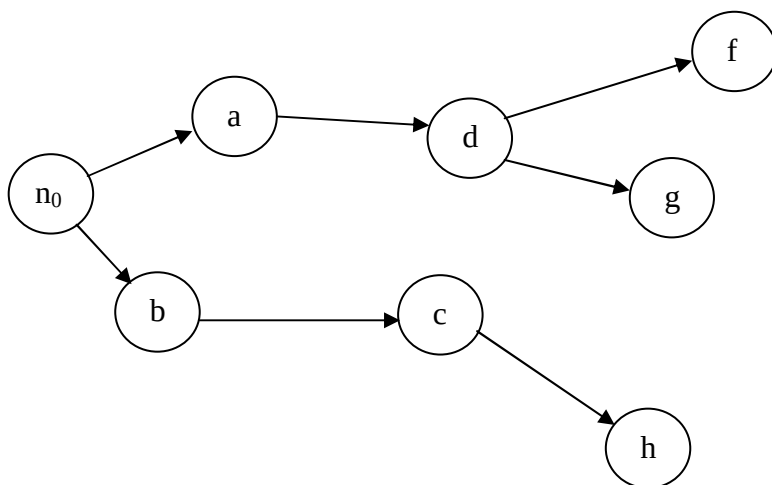
Între nodurile A și C și nodurile A și B avem câte două arce

#### 4.1.4. Arbori

O clasă specială de grafuri orientate sunt *arborii*.

Un graf orientat se numește *arbore* dacă avem un nod distinct numit *rădăcina arborelui*, notat cu  $n_0$ , astfel încât pentru orice nod din restul arborelui să avem un singur drum care să unească acel nod cu  $n_0$ , nodul-rădăcină.

Fie următorul arbore:

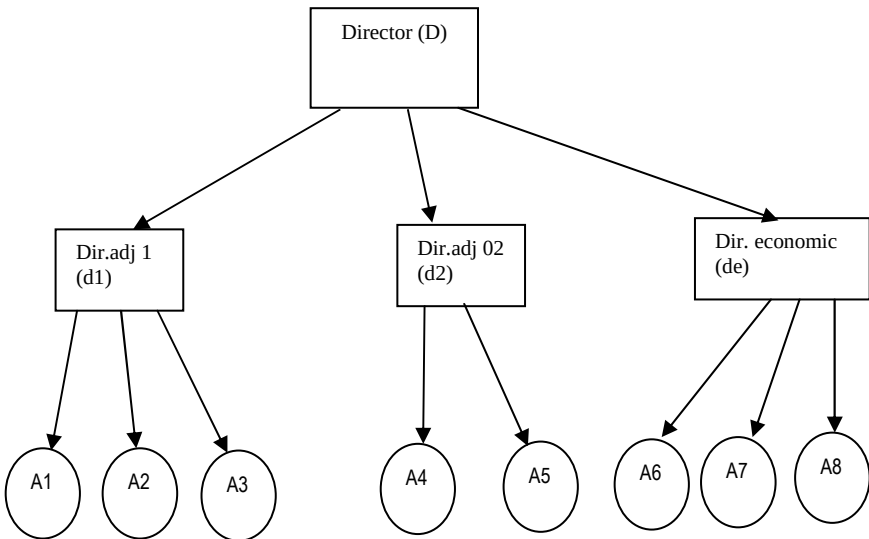


De exemplu, de la nodul  $n_0$  la nodul g avem drumul:

$$n_0 - a - d - g$$

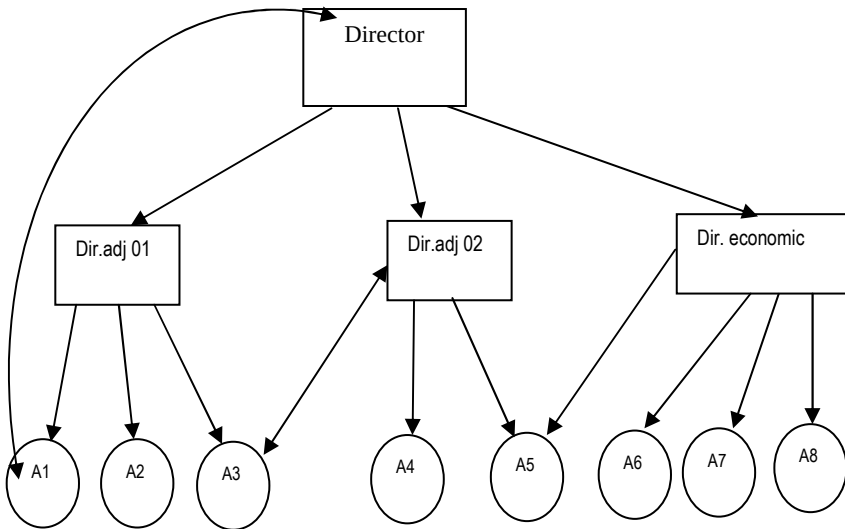
### Exemplul 79

Structura unei organizații poate fi pusă sub forma unui arbore:



Pornind de la acest exemplu putem stabili două tipuri de organizații:

- cele de tip arbore, în care între orice angajat și director există un drum unic, așa cum se vede în figura de mai sus.
- cele de tip graf multiplu în care între angajat și director sunt mai multe drumuri ca-n figura de mai jos:



De exemplu, între director și angajatul A5 avem două drumuri: director – dir. adj. 02 – A5 și director – dir economic – A5.

#### 4.1.4. Proprietăți ale grafurilor

Spunem că o relație este *tranzitivă* dacă din faptul că A este în relație cu B și B este în relație cu C rezultă că A este în relație cu C. Aceasta se mai scrie astfel:

$$A \text{ (rel)} B, B \text{ (rel)} C \rightarrow A \text{ (rel)} C$$

#### **Exemplu 79 (continuare)**

Să presupunem că directorul ia o decizie pe care o comunică directorului economic, deci  $D \text{ (rel)} de$ . Apoi directorul economic spune angajatului A7 să se ocupe de ducerea la îndeplinire a deciziei directorului, deci  $de \text{ (rel)} A7$ . În concluzie vom spune că decizia directorului a fost dusă la îndeplinire de

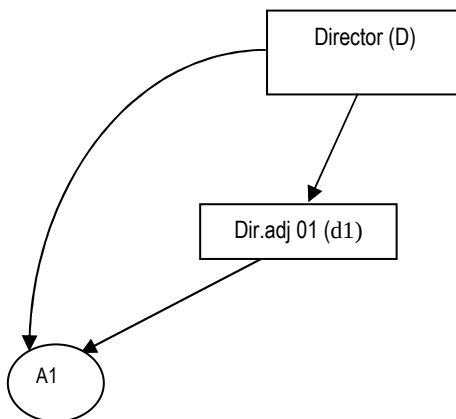
| angajatul A7, adică  $D(\text{rel})A7$ .

Un alt concept al teoriei grafurilor este cel de *distanță* dintre două noduri.

Pentru început fie un graf simplu. Atunci numărul de arce dintre două noduri îl definim ca fiind distanța dintre cele două noduri. Trecând la grafurile multiple se pune problema cum definim distanța dintre două noduri, având în vedere că, în acest caz, între două noduri am putea avea mai multe drumuri. Definim distanța dintre două noduri, notată cu  $d(x,y)$ , ca fiind cel mai scurt drum între nodurile  $x$  și  $y$ . Deoarece cazul grafului simplu reprezintă o particularizare a grafului multiplu putem generaliza definiția drumului pentru orice tip de graf.

### Exemplul 80

Fie structura unei organizații de tip graf multiplu, ca-n partea a doua a exemplului 4.2. Atunci distanța dintre director și A2 este egală cu 2 ( $d(D,A2) = 2$ ). Dar distanța dintre director și A1 este egală cu 1 pentru că, în această situație, avem două drumuri, unul format din două arce și unul format dintr-un singur arc.



Dacă între două noduri nu există niciun arc, atunci distanța dintre ele se va considera egală cu

infinit.

### **Exemplul 32 (continuare)**

În exemplul cu scrierea unei sociograme sub forma unui graf, se poate vedea că, de exemplu, distanța dintre S5 și S9 este egală cu 4 în timp ce distanța dintre S5 și S1 este  $\infty$ .

În cazul în care distanța dintre două noduri este infinit spunem că un nod este inabordabil pentru celălalt. În cazul de mai sus, vom spune că S1 este inabordabil pentru S5.

#### *4.1.6. Altă posibilitate de studiu al relațiilor dintr-un grup de indivizi*

În continuare, vom studia anumite relații dintr-un grup pornind de la conceptul matematic de *relație de echivalență*. Pentru început să reamintim ce înseamnă o relație de echivalență:

Spunem că **rel** este o relație de echivalență peste o mulțime M dacă:

- a.  $A \text{ rel } A$  pentru orice A din mulțimea M (reflexivitatea)
- b. Dacă  $A \text{ rel } B$  atunci și  $B \text{ rel } A$  pentru orice A și B din mulțimea M (simetria)
- c. Dacă  $A \text{ rel } B$  și  $B \text{ rel } C$  atunci  $A \text{ rel } C$  (tranzitivitatea)

Mulțimea între ale cărei elemente există o relație de echivalență se numește relație de echivalență.

În continuare analizăm relațiile dintr-un grup în funcție de cei care execută sarcinile. Definim o relație a unui grup astfel: spunem că A se află în relație cu B dacă A ia o decizie și B o execută. Afirmăm că legăturile într-un grup sunt cu atât mai puternice cu cât relația definită mai sus este mai aproape de o relație de echivalență.

**Reflexivitatea:** în acel grup fiecare are anumite sarcini de care răspunde personal, adică are de executat sarcini în care el ia decizia și tot el execută, adică  $A \text{ rel } A$

**Simetria:** Dacă pentru o parte din sarcini A răspunde și B trebuie să execute ceea ce a decis A, există sarcini în care B decide și A execută. De exemplu, în execuția unui proiect mai mare, într-o anumită etapă A poate fi specialistul, iar în altă etapă B poate fi specialistul.

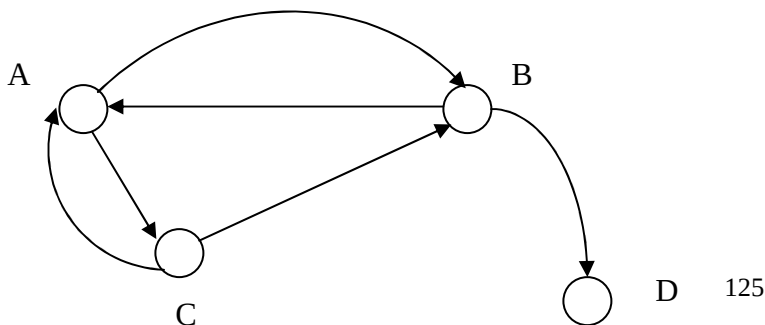
**Tranzitivitatea:** A decide un anumit lucru, B își dă seama că nu poate finaliza sarcina dată de A și delegă pe C să finalizeze. Deci, în continuare, la recomandarea lui B A va lua decizii, iar C le va executa.

Desigur, toate acestea se petrec într-un caz ideal. Dar totuși se pot găsi niște cuantificatori prin care se poate măsura „apropierea” de o relație de echivalență astfel definită. Bineînțeles că acesta e doar un exemplu și ar putea exista cuantificări mai bune a relațiilor dintr-un anumit grup.

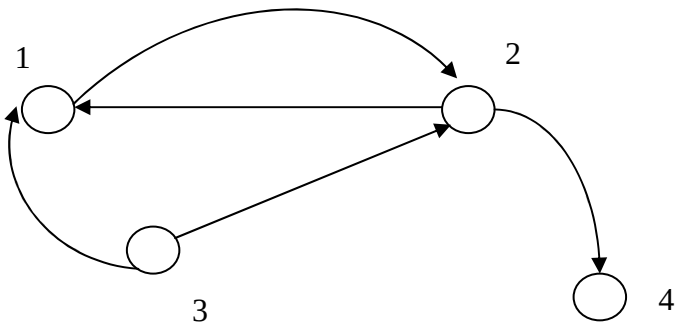
#### 4.1.7. Reprezentarea matriceală a unui graf

Deoarece grafurile nu sunt doar un mod de reprezentare a unor relații între anumiți indivizi, ci ele pot reprezenta un instrument util în modelarea matematică și având vedere că astăzi orice modelare înseamnă și introducerea datelor pe calculator, se pune problema cum se pot scrie grafurile astfel încât să fie cât mai ușor citite de un program pe calculator. Pentru o înțelegere mai bună vom lua un exemplu:

Fie un graf reprezentat ca mai jos:



În primul rând vom realiza o numerotare a nodurilor. Atunci graful va deveni:



Atunci putem avea următorul tabel în care pe linie punem nodul de plecare și pe coloană punem nodul în care ajunge arcul. În căsuțele tabelului vom pune 1 dacă avem un arc între noduri și 0 dacă nu există un arc între noduri. Astfel, în căsuța corespunzătoare liniei a doua și coloana întâi am pus valoarea 1 pentru că există un arc de la nodul 2 la nodul 1. Folosind aceeași regulă, în linia a 3-a și coloana a 4-a am pus valoarea 0 pentru că nu există niciun arc direct între nodurile 3 și 4. În urma completării obținem următorul tabel:

|   | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 0 | 1 | 0 | 0 |
| 2 | 1 | 0 | 0 | 1 |
| 3 | 1 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 |

Dacă eliminăm acum prima linie și prima coloană, corespunzătoare etichetelor, vom obține matricea asociată grafului:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Această matrice este ușor de folosit pentru realizarea unui program pe calculator.

#### 4.2. Baze de date

Fie imediat după culegerea de pe teren, fie între etapele de analiză, datele trebuie stocate, astfel încât să poată fi accesate cât mai ușor. De aceea, a apărut ca o etapă importantă a cercetării stocarea datelor într-o bază de date. Ce este o bază de date? Putem da o definiție a bazei de date în funcție de facilitățile pe care le oferă: *Locul unde se pun datele astfel încât să le putem înregistra cu toate informațiile specifice disponibile, să le putem căuta, să le putem extrage după anumite criterii cerute într-o anumită etapă a cercetării.*

O bază de date poate fi construită ca un tabel în care liniile se numesc *înregistrări*, iar coloanele se numesc *câmpuri*.

#### Exemplul 81

Fie o bază de date cu o baterie de 3 teste aplicate la 10 subiecți. Aceasta poate fi reprezentată ca un tabel astfel:

|                   | <i>Test 1</i> | <i>Test 2</i> | <i>Test 3</i> |
|-------------------|---------------|---------------|---------------|
| <i>Subiect 1</i>  | 12            | 2             | 1             |
| <i>Subiect 2</i>  | 11            | 5             | 1             |
| <i>Subiect 3</i>  | 14            | 4             | 4             |
| <i>Subiect 4</i>  | 15            | 1             | 1             |
| <i>Subiect 5</i>  | 16            | 5             | 1             |
| <i>Subiect 6</i>  | 17            | 2             | 2             |
| <i>Subiect 7</i>  | 12            | 2             | 1             |
| <i>Subiect 8</i>  | 10            | 4             | 2             |
| <i>Subiect 9</i>  | 11            | 6             | 1             |
| <i>Subiect 10</i> | 15            | 4             | 1             |

În acest caz prin înregistrare se înțeleg toate informațiile disponibile despre un subiect. De exemplu, înregistrarea corespunzătoare Subiectului 3 este

|                  |    |   |   |
|------------------|----|---|---|
| <i>Subiect 3</i> | 14 | 4 | 4 |
|------------------|----|---|---|

Un câmp este dat de valorile corespunzătoare unui test. De exemplu, avem câmpul Test 1:

|               |
|---------------|
| <i>Test 1</i> |
| 12            |
| 11            |
| 14            |
| 15            |
| 16            |
| 17            |
| 12            |
| 10            |
| 11            |
| 15            |

Bazele de date se pot clasifica în baze de date analitice și baze de date operaționale.

*Bazele de date analitice* sunt bazele de date de tip static în care datele sunt sub forma unei liste. Problema constă în căutarea în această bază de date care, în cazul unor baze de date foarte mari, poate deveni foarte anevoioasă.

Odată cu apariția computerelor a apărut un al doilea tip: *bazele de date operaționale*. Acest gen de baze de date permite ceva mai mult decât o simplă vizualizare a datelor. Putem realiza o serie de operații, cum ar fi: adăugarea de înregistrări, modificarea înregistrărilor, sortarea după un anumit criteriu, ștergerea de înregistrări. Dintre soft-urile în care se pot crea baze de date operaționale amintim: Excel, Microsoft Acces, SPSS.

## 5. UTILIZAREA EXCEL ȘI SPSS ÎN STATISTICĂ

În acest capitol ne-am propus să facem o prezentare succintă a principalelor tehnici de utilizare a Excel-ului și a SPSS în calcularea unor indicatori și teste statistice. Desigur, nu ne-am propus să facem o prezentare exhaustivă a posibilităților acestor două soft-uri în calculele statistice. Pentru cei interesați în a aprofunda aceste două lucruri, recomandăm pentru Excel, de exemplu [Pom], iar pentru SPSS [How].

Acest capitol va fi împărțit în două: în prima parte vom prezenta utilizarea Excel-ului în statistică, iar în partea a doua utilizarea SPSS-ului în statistică. Prezentarea se va realiza pornind de la exemple concrete. La anumite exemple am modificat, în mod intenționat, datele pentru a evidenția anumite aspecte ale utilizării acestor soft-uri.

### 5.1. Utilizarea Excel în statistică

Excel-ul reprezintă un produs Microsoft, dezvoltat pentru sistemele Windows și Macintosh. Un fișier Excel cuprinde mai multe foi de calcul. Calculele se pot realiza pe o foaie de calcul, dar se pot realiza legături între foile de lucru ale aceluiași fișier, sau legături cu foi de lucru din alte fișiere. Amintim că, în afară de Excel, mai există și alte soft-uri ce pot realiza sarcini asemănătoare cu ale Excel-ului. De exemplu, din pachetul Open Office, avem Open Office Calc. Am optat totuși pentru prezentarea Excel-ului deoarece, la ora actuală, este cel mai răspândit, și, în general, un utilizator al Excel-ului se poate acomoda relativ rapid cu alte programe de calcul ce au apărut sau ar putea apărea pe piața soft-urilor.

În continuare vom considera cunoscut în Excel:

- deschiderea, salvarea unui fișier
- elementele meniului

- modul în care se specifică o anumită celulă
- introducerea datelor pe o foaie
- realizarea celor patru operații matematice între valorile cuprinse în anumite celule din foaia de calcul
- copierea formulelor.

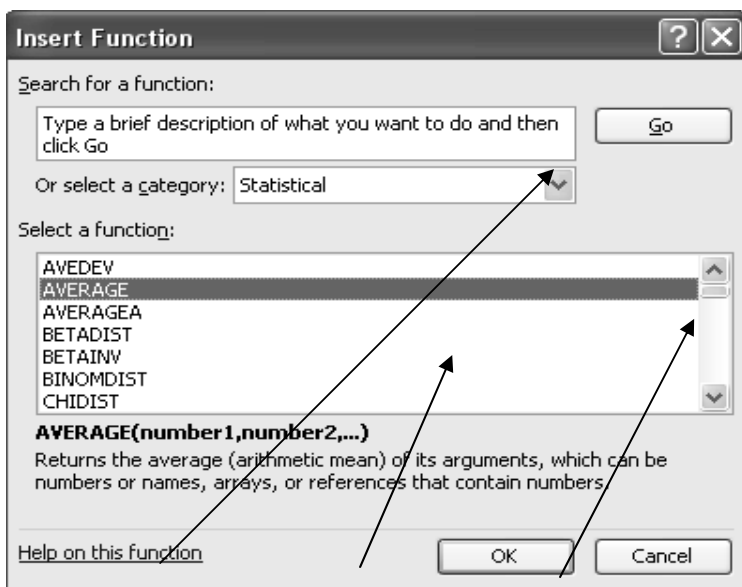
### 5.1.1. Inserarea unei funcții

Înainte de a trece la prezentarea funcțiilor specifice Excel-ului, vom reaminti modul în care se poate insera o funcție în Excel.

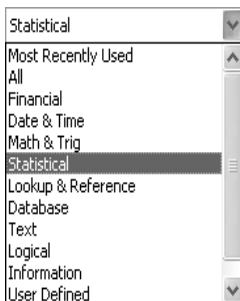
Pentru a insera o funcție dăm un clic pe butonul din meniu ce arată ca mai jos:



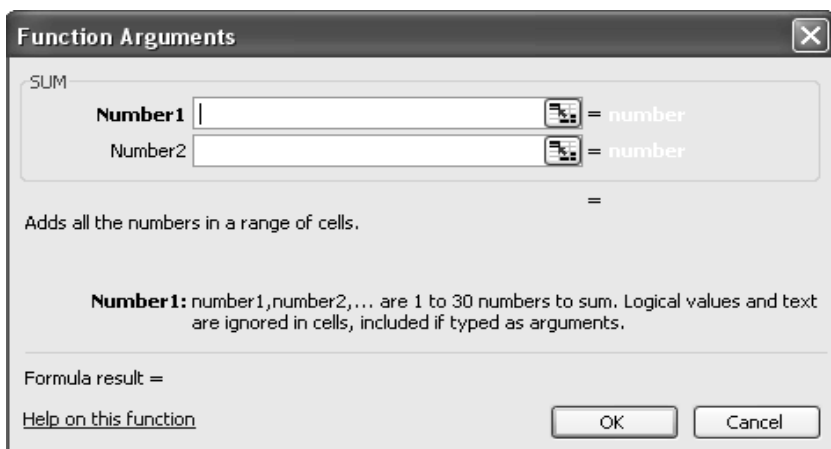
După ce am apăsât pe acest buton se va deschide o fereastră de forma următoare:



Dacă dăm clic pe butonul indicat de săgeata 1 se va deschide o subfereastră

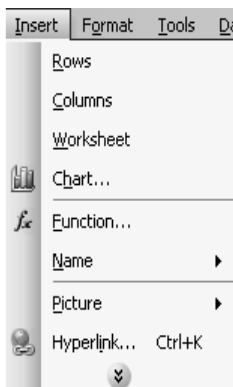


din care putem selecta tipul de funcție dorită (în situația de mai sus am selectat opțiunea **Statistical**). După selecție în fereastra indicată de săgeata 2 va apărea lista cu funcțiile corespunzătoare celui tip de funcție. De exemplu, pentru a selecta funcția SUM (se va prezenta mai jos această funcție) alegem categoria **Math & Trig** și apoi activăm fereastra 2 și dăm un clic pe butonul de jos al ruler-ului (marcat cu 3 în figura de mai sus), până se va vedea în listă denumirea SUM. Dăm un clic pe denumire și apoi apăsăm butonul OK. Se va obține fereastra



În primul dreptunghi punem celulele care vor intra în sumă.

De multe ori s-ar putea însă ca butonul  să nu fie pus în meniu. Atunci pentru a insera o funcție dăm un clic pe opțiunea **Insert** din meniu și din fereastra



Alegem opțiunea **Function**

În cele ce urmează vom relua exemplele din capitolul 3 și vom arăta cum se pot efectua calculele pentru a obține rezultatele prezentate în acele exemple.

#### 5.1.2. *Minim și maxim dintr-un șir de date*

##### **Exemplul 49 (continuare)**

Într-un liceu s-a studiat manifestarea complexului de inferioritate la un eșantion de 30 de adolescenți. S-a studiat posibilitatea ca să avem o manifestare a complexului de inferioritate prin valențe ridicate ale anxietății<sup>94</sup>. În urma aplicării testului s-au obținut următoarele rezultate:

---

<sup>94</sup> Am luat în considerare doar anxietatea manifestă.

|    | A          | B             | C          | D             | E |
|----|------------|---------------|------------|---------------|---|
| 1  |            | <b>Nota A</b> |            | <b>Nota A</b> |   |
| 2  | <b>S1</b>  | 32            | <b>S16</b> | 32            |   |
| 3  | <b>S2</b>  | 31            | <b>S17</b> | 31            |   |
| 4  | <b>S3</b>  | 33            | <b>S18</b> | 33            |   |
| 5  | <b>S4</b>  | 27            | <b>S19</b> | 27            |   |
| 6  | <b>S5</b>  | 29            | <b>S20</b> | 29            |   |
| 7  | <b>S6</b>  | 25            | <b>S21</b> | 25            |   |
| 8  | <b>S7</b>  | 23            | <b>S22</b> | 23            |   |
| 9  | <b>S8</b>  | 22            | <b>S23</b> | 22            |   |
| 10 | <b>S9</b>  | 25            | <b>S24</b> | 25            |   |
| 11 | <b>S10</b> | 21            | <b>S25</b> | 21            |   |
| 12 | <b>S11</b> | 25            | <b>S26</b> | 25            |   |
| 13 | <b>S12</b> | 25            | <b>S27</b> | 25            |   |
| 14 | <b>S13</b> | 19            | <b>S28</b> | 19            |   |
| 15 | <b>S14</b> | 23            | <b>S29</b> | 23            |   |
| 16 | <b>S15</b> | 27            | <b>S30</b> | 27            |   |
| 17 |            |               |            |               |   |
| 18 | minim      |               |            |               |   |
| 19 | maxim      |               |            |               |   |
| 20 |            |               |            |               |   |

Pentru a calcula minimul și apoi maximul acestor date, mai întâi le vom pune într-o foaie de Excel. Dacă le-am aranja pe o singură coloană, s-ar putea să nu putem vizualiza pe ecran toate datele. O soluție ar fi să punem datele în Excel exact ca-n tabelul de mai sus.

Atunci când vom selecta tot tabelul și vom calcula, cu ajutorul funcțiilor maximul și minimul acestor date, datele de tip caracter vor fi ignorate. În celula B18 vom calcula minimul folosind formula „=MIN (A2:D16)”. Apoi, apăsând pe *Enter*, vom obține în B18 valoarea **19**. Asemănător, se va calcula și valoarea maximă a datelor, folosind în celula B19 formula „=MAX (A2:D16)”. Maximul obținut va fi **33**.

### 5.1.3. Media

#### **Exemplul 50 (continuare)**

|    | A  | B           | C            | D |
|----|----|-------------|--------------|---|
| 1  |    | mem vizuala | mem auditiva |   |
| 2  | LM | 5           | 6            |   |
| 3  | VA | 3           | 4            |   |
| 4  | MA | 2           | 4            |   |
| 5  | TC | 4           | 6            |   |
| 6  | SA | 2           | 2            |   |
| 7  | TS | 2           | 3            |   |
| 8  | BR | 4           | 5            |   |
| 9  | FI | 4           | 4            |   |
| 10 | DC | 6           | 8            |   |
| 11 | CR | 1           | 2            |   |
| 12 | LL | 3           | 4            |   |
| 13 |    |             |              |   |

În urma aplicării unui test de memorie pe un lot de copii s-au obținut următoarele rezultate (datele au fost deja trecute pe o foaie de calcul din Excel)

Pentru a calcula media pentru memoria vizuală și memoria auditivă vom folosi formula AVERAGE. De exemplu, pentru a calcula media valorilor memoriei vizuale vom scrie în celula B13: „=AVERAGE (B2:B12)”. Rezultatul va fi 3,273.

#### **Exemplul 53 (continuare)**

S-a aplicat un test de atenție concentrată la copii din clasa a III-a. În urma testului s-au obținut următoarele rezultate:

|   | A                             | B       | C |
|---|-------------------------------|---------|---|
| 1 | Atenție concentrată-cantitate |         |   |
| 2 | Valori                        | Ponderi |   |
| 3 | 60                            | 1       |   |
| 4 | 70                            | 2       |   |
| 5 | 80                            | 1       |   |
| 6 | 90                            | 11      |   |
| 7 | 100                           | 7       |   |
| 8 |                               |         |   |

Reamintim că ponderile reprezintă numărul de subiecți care au valoarea corespunzătoare din stânga. De exemplu, cu valoarea

70 avem 2 subiecți. Pentru a calcula media, în acest caz, trebuie să înmulțim respectiv elementele din fiecare coloană. Pentru această operație avem funcția SUMPRODUCT. Pentru a calcula numărul de elemente dintr-un șir de numere vom folosi funcția COUNT. Deci, pentru a calcula media ponderată pentru datele din exemplu vom scrie în celula A8 următoarea formulă: „=SUMPRODUCT (A3:A7,B3:B7)/SUM (B3:B7)”. Rezultatul obținut va fi: 89,545

#### 5.1.4. Mediana

##### **Exemplul 54 (continuare)**

Studentii au vrut să-și aleagă un șef de grupă. Pentru că nu au ajuns la un consens, au rugat un profesor să-i ajute în această problemă. În urma discuțiilor avute cu studenții profesorul a determinat 3 candidați pentru această funcție. Apoi a pus pe fiecare student să acorde punctaje între 1 și 5 celor 3 candidați.

S-au obținut datele de mai sus

|    | A       | B   | C   | D   | E |
|----|---------|-----|-----|-----|---|
| 1  |         | S10 | S15 | S22 |   |
| 2  | S1      | 1   | 3   | 5   |   |
| 3  | S2      | 2   | 3   | 5   |   |
| 4  | S3      | 1   | 3   | 5   |   |
| 5  | S4      | 2   | 4   | 5   |   |
| 6  | S5      | 5   | 3   | 2   |   |
| 7  | S6      | 1   | 4   | 3   |   |
| 8  | S7      | 5   | 3   | 2   |   |
| 9  | S8      | 2   | 4   | 1   |   |
| 10 | S9      | 2   | 3   | 2   |   |
| 11 | S10     |     | 1   | 2   |   |
| 12 | S11     | 5   | 3   | 2   |   |
| 13 | S12     | 5   | 3   | 1   |   |
| 14 | S13     | 5   | 4   | 2   |   |
| 15 | S14     | 2   | 3   | 5   |   |
| 16 | S15     | 5   |     | 5   |   |
| 17 | S16     | 5   | 4   | 2   |   |
| 18 | S17     | 2   | 3   | 2   |   |
| 19 | S18     | 5   | 3   | 5   |   |
| 20 | S19     | 5   | 3   | 2   |   |
| 21 | S20     | 2   | 3   | 5   |   |
| 22 | S21     | 5   | 4   | 5   |   |
| 23 | S22     | 2   | 3   |     |   |
| 24 |         |     |     |     |   |
| 25 | mediana |     |     |     |   |
| 26 |         |     |     |     |   |

Pentru a calcula mediana corespunzătoare valorilor obținute de S10 vom scrie în B25 următoarea formulă: „=MEDIAN (B2:B23)”. Rezultatul va fi 2. Asemănător se va obține pentru S15 valoarea medianei egală cu 3, respectiv pentru S22 o valoare 2.

### Exemplele 55 și 56 (continuare)

La două echipe de muncitori, echipa A și echipa B, li s-au cerut să noteze cu note între 1 și 10 un șef de echipă ce a lucrat cu amândouă echipele. În urma centralizării s-au obținut următoarele rezultate:

|   | A             | B   | C   | D   | E   | F   | G       | H       | I |
|---|---------------|-----|-----|-----|-----|-----|---------|---------|---|
| 1 |               | AM1 | AM2 | AM3 | AM4 | AM5 | mediana |         |   |
| 2 | Note acordate | 2   | 2   | 3   | 4   | 10  |         |         |   |
| 3 |               |     |     |     |     |     |         |         |   |
| 4 |               | BM1 | BM2 | BM3 | BM4 | BM5 | BM6     | mediana |   |
| 5 | Note acordate | 1   | 1   | 2   | 3   | 3   | 3       |         |   |
| 6 |               |     |     |     |     |     |         |         |   |
| 7 |               |     |     |     |     |     |         |         |   |

În celula G3 se calculează mediana cu formula: „=MEDIAN (B2:F2)”. Rezultatul va fi 3. Asemănător se calculează mediana în H5 și se obține valoarea 2,5.

#### 5.1.5. Amplitudinea

### Exemplu 49 (continuare)

Reluăm exemplul 5.1. Dorim pentru subiecții de mai jos să calculăm amplitudinea.

Pentru aceasta în D18 vom scrie formula:

=MAX (B2:D16)-MIN (B2:B16)

Dând *Enter* vom obține rezultatul 14.

### 5.1.6. Coeficientul de omogenitate

#### **Exemplu 64 (continuare)**

Pe un lot de studenți în sesiune s-a măsurat anxietatea voalată dimineața și seara. În urma centralizării datelor într-o foaie de Excel s-a obținut următorul tabel:

|    | A       | B         | C     | D |
|----|---------|-----------|-------|---|
| 1  |         | dimineata | seara |   |
| 2  |         | 21        | 20    |   |
| 3  |         | 16        | 17    |   |
| 4  |         | 15        | 16    |   |
| 5  |         | 24        | 22    |   |
| 6  |         | 16        | 18    |   |
| 7  |         | 12        | 18    |   |
| 8  |         | 23        | 23    |   |
| 9  |         | 15        | 16    |   |
| 10 |         | 16        | 17    |   |
| 11 |         | 18        | 19    |   |
| 12 |         | 24        | 26    |   |
| 13 |         | 21        | 20    |   |
| 14 |         | 13        | 15    |   |
| 15 |         | 23        | 22    |   |
| 16 |         | 16        | 17    |   |
| 17 |         | 14        | 16    |   |
| 18 |         | 14        | 15    |   |
| 19 |         | 12        | 14    |   |
| 20 |         | 19        | 18    |   |
| 21 |         | 15        | 16    |   |
| 22 | medie   |           |       |   |
| 23 | abt std |           |       |   |
| 24 | Cv      |           |       |   |
| 25 |         |           |       |   |

În primul rând vom calcula media pentru cele două șiruri de date în celulele B22 și C22 cu ajutorul funcției AVERAGE și pentru primul lot vom obține valoarea 17,35, iar pentru al doilea lot vom obține valoarea 18,25. Pentru a calcula abaterea standard vom folosi formula „=STDEV (B2:B21)” pentru primul lot și „=STDEV (C2:C21)” pentru al doilea lot. Pentru primul lot vom obține valoarea 4,00, iar

pentru al doilea lot vom obține valoarea 3,09. Pentru a calcula coeficientul de omogenitate va trebui să calculăm raportul dintre abaterea standard și medie și apoi să-l exprimăm în procente. Pentru raport, în celula B24 vom scrie: „=B23/B22” și vom obține valoarea 0,2307. Pentru a o transforma în procent dăm un clic pe celula B24, apoi un clic pe butonul



din meniu. Pentru a obține un procent cu o

zecimală vom da mai apoi și un clic pe butonul  din meniu. Rezultatul final pentru primul lot va fi 23,1%. Asemănător se va proceda și pentru al doilea lot pentru care vom avea rezultatul 16,9%.

#### 5.1.7. Compararea mediei unui grup cu o valoare dată

Deoarece în capitolul 3, la exemplu 68 nu sunt date toate datele vom da un alt exemplu, pornind de la exemplul 64.

#### **Exemplul 64** (date modificate)

Pe un lot de studenți în sesiune s-a studiat anxietatea voalată seara. În urma centralizării datelor s-a obținut următorul tabel:

|    | A          | B     | C |
|----|------------|-------|---|
| 1  |            | seara |   |
| 2  |            | 20    |   |
| 3  |            | 17    |   |
| 4  |            | 16    |   |
| 5  |            | 22    |   |
| 6  |            | 18    |   |
| 7  |            | 18    |   |
| 8  |            | 22    |   |
| 9  |            | 17    |   |
| 10 |            | 16    |   |
| 11 |            | 15    |   |
| 12 |            | 14    |   |
| 13 |            | 18    |   |
| 14 |            | 16    |   |
| 15 | medie      | 17.62 |   |
| 16 | valoarea z |       |   |
| 17 |            |       |   |

Știind din alte cercetări că, în mod normal, media studenților în sesiune este de 19 se cere să studieze dacă valoarea obținută pe lotul studiat e diferită de valoarea populației de studenți obținută în studiile anterioare.

Pentru a rezolva această problemă vom folosi funcția ZTEST. Dăm un clic în celula B16, apoi dăm un clic pe butonul de funcții și din listă alegem funcția ZTEST. Se va deschide o fereastră ca mai jos:

Function Arguments

ZTEST

Array  = number

X  = number

Sigma  = number

Returns the two-tailed P-value of a z-test.

Array is the array or range of data against which to test X.

Formula result =

[Help on this function](#)

OK Cancel

În dreptunghiul **Array** vom scrie șirul de date: B2:B14, în dreptul lui **X** se va pune valoarea mediei populației. Deoarece nu cunoaștem valoarea dispersiei populației de studenți nu vom completa nimic la **Sigma** și atunci se va lua valoarea dispersiei lotului de subiecți. După ce s-au completat spațiile libere obligatorii, se va da un clic pe butonul . Valoarea obținută va fi 0,978.

### 5.1.8. Testul $t$

#### Exemplul 59

În urma aplicării testului Guilford de abilități ale gândirii divergente (flexibilitatea), la un liceu s-au obținut următoarele rezultate:

|    | A       | B  | C    | D  | E  | F |
|----|---------|----|------|----|----|---|
| 1  | Băieți  |    | Fete |    |    |   |
| 2  | 5       | 7  | 8    | 6  | 14 |   |
| 3  | 11      | 5  | 5    | 5  | 7  |   |
| 4  | 11      | 11 | 12   | 8  | 7  |   |
| 5  | 5       | 7  | 7    | 6  | 9  |   |
| 6  | 7       | 7  | 11   | 3  | 9  |   |
| 7  | 6       | 7  | 5    | 6  |    |   |
| 8  | 13      | 8  | 7    | 8  |    |   |
| 9  | 6       | 12 | 5    | 6  |    |   |
| 10 | 8       | 12 | 8    | 7  |    |   |
| 11 | 16      | 10 | 10   | 16 |    |   |
| 12 | 7       | 11 | 14   | 12 |    |   |
| 13 | 6       | 7  | 7    | 6  |    |   |
| 14 |         |    |      |    |    |   |
| 15 | eroarea |    |      |    |    |   |
| 16 |         |    |      |    |    |   |

În Excel rezultatul funcției TTEST este eroarea (P-value). Iar pentru a avea diferențe semnificative trebuie ca valoarea rezultată în urma aplicării funcției să fie mai mică decât 0,05.

În primul rând vom da un clic pe celula B15. Selectăm din lista de funcții statistice, funcția TTEST și va apărea următoarea fereastră:

**Function Arguments**

TTEST

|               |                      |          |
|---------------|----------------------|----------|
| <b>Array1</b> | <input type="text"/> | = array  |
| <b>Array2</b> | <input type="text"/> | = array  |
| <b>Tails</b>  | <input type="text"/> | = number |
| <b>Type</b>   | <input type="text"/> | = number |

=

Returns the probability associated with a Student's t-Test.

**Array1** is the first data set.

Formula result =

[Help on this function](#)

OK Cancel

Dăm un clic pe dreptunghiul din dreptul etichetei **Array1**. Scriem acolo celulele în care se găsește primul șir de numere<sup>95</sup>: A2:B13. Analog în dreptul etichetei **Array2** scriem<sup>96</sup>: C2:E13. La opțiunea **Tails** se pune 1 dacă am făcut o presupunere privitor la poziția unei medii față de cealaltă (de exemplu, media fetelor este mai mare decât media obținută de băieți), sau punem valoarea 2 dacă ipoteza era că cele două medii sunt diferite fără însă a specifica ce medie este mai mare.

Pentru opțiunea **Type** avem trei posibilități.

- 1 – dacă șirurile comparate reprezintă fizic același șir studiat înainte de a aplica un stimul și după ce am aplicat un stimul. De exemplu, dacă studiem depresia la un lot de pacienți înainte de operație și după operație, fizic lotul este același și ne interesează să vedem dacă stimulul

<sup>95</sup> În exemplul prezentat intenționat am pus șirul pe două coloane, pentru a arăta că nu este obligatoriu ca șirul să fie pe o singură coloană.

<sup>96</sup> Aici se observă că în selecția făcută pot fi și celule goale, acestea fiind ignorate de către funcția TTEST.

(în acest caz operația) a avut vreo influență asupra pacienților

- 2 – dacă șirurile au aceeași dispersie
- 3 – dacă șirurile au dispersie diferită

În cazul prezentat la Tails vom alege opțiunea 2, iar la opțiunea Type alegem opțiunea 3. Rezultatul final va fi: 0,573, deci eroarea este mai mare de 0,05, adică nu avem diferențe semnificative între cele două medii. Se observă că pentru folosirea funcției TTEST nu avem nevoie să calculăm mediile celor două loturi. Acestea trebuie calculate separat atunci când, în cazul diferențelor semnificative, dorim să reprezentăm grafic mediile comparate.

### 5.1.9. Testul $\chi^2$

#### **Exemplul 72 (continuare)**

La un sondaj realizat în 2004 pe un eșantion de 1500 persoane (subiecți), printre altele, s-au adresat următoarele întrebări cu variantele de răspuns:

3. Câtă încredere aveți în Biserică?
  - a. Foarte multă
  - b. Multă
  - c. Puțină
  - d. Foarte puțină
  - e. Deloc
4. Câtă încredere aveți în armată?
  - a. Foarte multă
  - b. Multă
  - c. Puțină
  - d. Foarte puțină
  - e. Deloc

În urma centralizării rezultatelor s-a obținut următorul tabel:

|   | A        | B            | C     | D      | E             | F     | G     | H |
|---|----------|--------------|-------|--------|---------------|-------|-------|---|
| 1 |          | Foarte multă | Multă | Puțină | Foarte puțină | Deloc | TOTAL |   |
| 2 | Biserica | 626          | 525   | 201    | 58            | 90    | 1500  |   |
| 3 | Armata   | 377          | 694   | 289    | 54            | 86    | 1500  |   |
| 4 |          |              |       |        |               |       |       |   |
| 5 | eroare   |              |       |        |               |       |       |   |
| 6 |          |              |       |        |               |       |       |   |
| 7 |          |              |       |        |               |       |       |   |
| 8 |          |              |       |        |               |       |       |   |
| 9 |          |              |       |        |               |       |       |   |

Pentru a vedea dacă avem diferențe semnificative între cele două distribuții vom folosi funcția CHITEST. Dăm un clic în celula B5 și apoi selectăm din lista cu funcții statistice funcția CHITEST. Va apărea următoarea fereastră:

Function Arguments

CHITEST

Actual\_range

= array

Expected\_range

= array

=

Returns the test for independence; the value from the chi-squared distribution for the statistic and the appropriate degrees of freedom.

Actual\_range

is the range of data that contains observations to test against expected values.

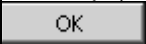
Formula result =

[Help on this function](#)

OK

Cancel

Dacă studiem faptul că a doua distribuție diferă semnificativ sau nu de prima distribuție, atunci la **Actual\_range** vom scrie B2:F2 iar la **Expected\_range** vom scrie B3:F3. (Reamintim că pentru a compara două distribuții vom lucra cu frecvențe absolute și suma frecvențelor de la prima distribuție trebuie să fie egală cu suma de la a doua distribuție). După ce vom da un clic pe butonul

 în celula B5 va apărea rezultatul 3.14879E-49. Aceasta este o scriere matematică a numărului zecimal subunitar ce are 48 de zecimale egale cu zero, iar a 49-a zecimală este 3 și a 50-a este 1 etc. Deci, este evident că numărul obținut este mai mic de 0,05, deci între cele două distribuții avem diferențe semnificative. O altă posibilitate de a analiza numărul obținut este să dăm un clic pe celula B5 și apoi dăm un clic pe butonul , și atunci va apărea în celula B5 valoarea 0%. Cum 0,05 se mai poate scrie 5% vom compara valoarea obținută cu ajutorul funcției cu 5%. Cum valoarea este mai mică decât 5%, putem spune că avem diferențe semnificative între distribuții.

#### 5.1.10. Coeficientul de corelație

##### **Exemplu 73** (Pearson și Lee)

S-a măsurat pe un lot de 11 familii statura fraților și surorilor în inci. După transformarea în cm se obține următorul tabel:

|   | A                       | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | L   | M |
|---|-------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|
| 1 | nr familiei             | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  |   |
| 2 | Sora                    | 175 | 163 | 165 | 160 | 165 | 157 | 165 | 163 | 168 | 150 | 157 |   |
| 3 | Frate                   | 180 | 173 | 168 | 170 | 178 | 180 | 178 | 179 | 183 | 165 | 168 |   |
| 4 |                         |     |     |     |     |     |     |     |     |     |     |     |   |
| 5 | coeficient de corelație |     |     |     |     |     |     |     |     |     |     |     |   |
| 6 |                         |     |     |     |     |     |     |     |     |     |     |     |   |
| 7 |                         |     |     |     |     |     |     |     |     |     |     |     |   |

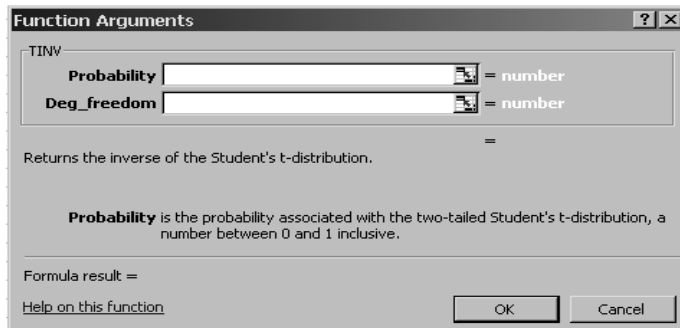
Să se studieze coeficientul de corelație între șirul de date corespunzător soriei și șirul de date corespunzător fratelui.

În acest caz vom folosi funcția CORREL. O vom selecta din lista funcțiilor statistice și vom obține fereastra:

La **Array1** vom scrie B2:L2, iar la **Array2** scrie B3:L3. Coeficientul de corelație va fi 0,6333. Pentru a verifica semnificabilitatea coeficientului de corelație va trebui să calculăm valoarea t asociată coeficientului. Reamintim formula:

$$t_{calc} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}$$

În B5 avem valoarea coeficientului de corelație, iar în C5 scriem numărul de înregistrări, adică valoarea 11. În B6 vom calcula valoarea lui t cu formula: „=B5\*sqrt (C5-2)/sqrt (1-B5^2)”. Vom obține valoarea 2,4547. Pentru a obține valoarea t din tabel vom folosi funcția TINV pe care o vom alege tot din lista funcțiilor statistice. Se va deschide fereastra:



În dreptul opțiunii **Probability** se va pune nivelul de semnificabilitate, adică 0,05<sup>97</sup>, iar în dreptul gradelor de libertate se va pune valoarea lui n-2, adică în cazul nostru 9<sup>98</sup>. După ce vom da un clic pe butonul OK, vom obține valoarea 2,2622. Deoarece valoarea calculată este mai mare decât valoarea din tabel rezultă că avem o corelație semnificativă între cele două șiruri de numere.

#### 5.1.11. Regresia liniară

##### **Exemplul 78**

Pe un grup de 20 de elevi s-a studiat corelația dintre memoria auditivă și mediile obținute pe semestrul 1. În urma centralizării datelor s-au obținut următoarele rezultate.

<sup>97</sup> Sau 5%.

<sup>98</sup> rezultat obținut din  $n = 11$  minus 2.

|    | A   | B                | C    | D |
|----|-----|------------------|------|---|
| 1  |     | Memoria auditiva | Note |   |
| 2  | S1  | 30               | 6.25 |   |
| 3  | S2  | 25               | 9.10 |   |
| 4  | S3  | 10               | 5.42 |   |
| 5  | S4  | 20               | 7.11 |   |
| 6  | S5  | 70               | 9.15 |   |
| 7  | S6  | 20               | 5.10 |   |
| 8  | S7  | 50               | 8.16 |   |
| 9  | S8  | 50               | 6.41 |   |
| 10 | S9  | -                | 5.30 |   |
| 11 | S10 | 20               | 5.32 |   |
| 12 | S11 | 0                | 6.33 |   |
| 13 | S12 | 50               | 9.15 |   |
| 14 | S13 | -                | 7.55 |   |
| 15 | S14 | 30               | 8.16 |   |
| 16 | S15 | 50               | 9.85 |   |
| 17 | S16 | 75               | 8.65 |   |
| 18 | S17 | -                | 4.65 |   |
| 19 | S18 | 0                | 6.35 |   |
| 20 | S19 | 10               | 5.45 |   |
| 21 | S20 | -                | 7.56 |   |
| 22 | S21 | 10               | 5.12 |   |
| 23 | S22 | 10               | 8.14 |   |
| 24 | S23 | -                | 8.15 |   |
| 25 | S24 | 20               | 4.20 |   |
| 26 | S25 | 10               | 6.48 |   |
| 27 |     |                  |      |   |

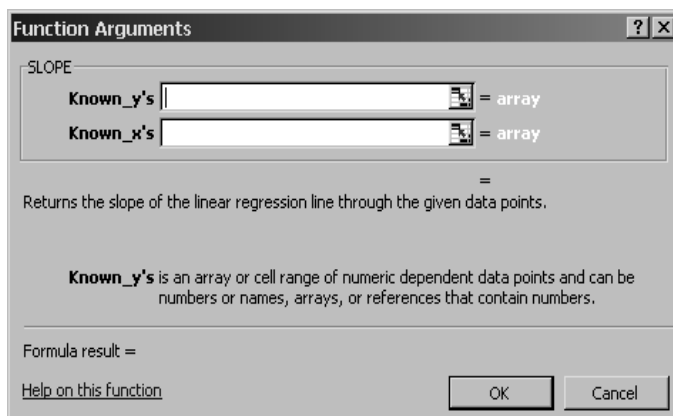
Se cere să se estimeze cele 5 rezultate lipsă.

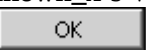
Pentru aceasta așezăm într-un tabel doar rezultatele corespunzătoare subiecților cărora li s-au aplicat testul de memorie auditivă.

|    | A     | B                  | C    | D |
|----|-------|--------------------|------|---|
| 1  |       | Memoria auditiva   | Note |   |
| 2  | S1    | 30                 | 6.25 |   |
| 3  | S2    | 25                 | 9.10 |   |
| 4  | S3    | 10                 | 5.42 |   |
| 5  | S4    | 20                 | 7.11 |   |
| 6  | S5    | 70                 | 9.15 |   |
| 7  | S6    | 20                 | 5.10 |   |
| 8  | S7    | 50                 | 8.16 |   |
| 9  | S8    | 50                 | 6.41 |   |
| 10 | S10   | 20                 | 5.32 |   |
| 11 | S11   | 0                  | 6.33 |   |
| 12 | S12   | 50                 | 9.15 |   |
| 13 | S14   | 30                 | 8.16 |   |
| 14 | S15   | 50                 | 9.85 |   |
| 15 | S16   | 75                 | 8.65 |   |
| 16 | S18   | 0                  | 6.35 |   |
| 17 | S19   | 10                 | 5.45 |   |
| 18 | S21   | 10                 | 5.12 |   |
| 19 | S22   | 10                 | 8.14 |   |
| 20 | S24   | 20                 | 4.20 |   |
| 21 | S25   | 10                 | 6.48 |   |
| 22 |       |                    |      |   |
| 23 |       | coef. de corelatie |      |   |
| 24 |       | 0.638              |      |   |
| 25 | val a |                    |      |   |
| 26 | val b |                    |      |   |
| 27 |       |                    |      |   |

Calculăm coeficientul de corelație pentru noile date și obținem valoarea 0,638. Făcând verificările, ca în exemplul anterior, rezultă că avem corelație semnificativă. Vom considera variabila *Memoria auditivă*, ca fiind variabilă dependentă și variabila *Note* ca fiind variabila independentă. Deci putem calcula regresia liniară de forma  $y = a \cdot x + b$ , unde  $x$  reprezintă variabila *Note*, iar  $y$  reprezintă variabila *Memoria auditivă*. Pentru a calcula pe  $a$ , numit și panta regresiei, vom folosi funcția SLOPE. Dăm un clic pe celula B25 și, după ce selectăm această funcție din lista funcțiilor

statistice, va apărea următoarea fereastră:



În conformitate cu convențiile stabilite anterior, la dreptunghiul din dreptul etichetei **Known\_y's** vom scrie B2:B21, iar la **Known\_x's** vom scrie C2:C21. Dăm un clic pe butonul  și vom obține în celula B25 valoarea 8,469. Pentru a afla valoarea lui b, termenul liber, vom folosi funcția INTERCEPT pe care o găsim tot în lista cu funcții statistice. Folosind un procedeu asemănător cu cel de la funcția SLOPE, vom obține valoarea -31,239. Pentru a estima valorile lipsă din tabel este suficient să calculăm pentru S9, în celula B10 din primul tabel valoarea expresiei: „=B25\*C10+B26” și vom obține valoarea estimată pentru S9 egală cu 13,645.

#### 5.1.12. Tabel funcții Excel pentru calcule statistice

Prin **Celinc** vom înțelege celula de la începutul șirului de date căruia dorim să-i aplicăm funcția, iar prin **Celfin** vom înțelege celula de la finalul șirului de date căruia dorim să-i aplicăm funcția

| <b><i>Denumire funcție</i></b> | <b><i>Forma funcției</i></b>                                        | <b><i>Unde se poate folosi funcția</i></b>                                   |
|--------------------------------|---------------------------------------------------------------------|------------------------------------------------------------------------------|
| Maxim                          | =MAX (Celinc:Celfin)                                                | Calcularea maximului unui șir de numere                                      |
| Minim                          | =MIN (Celinc:Celfin)                                                | Calcularea minimului unui șir de numere                                      |
| Media                          | =AVERAGE (Celinc:Celfin)                                            | Calculează media unui șir de numere                                          |
| Media ponderată                | =SUMPRODUCT (Celinc1:Celfin1,Celinc2:Celfin2)/SUM (Celinc2:Celfin2) | Calculează media ponderată                                                   |
| Mediana                        | =MEDIAN (Celinc:Celfin)                                             | Calculează mediana unui șir de numere                                        |
| Amplitudinea                   | =MAX (Celinc:Celfin)-MIN (Celinc:Celfin)                            | Calculează amplitudinea unui șir de date                                     |
| Coeficient de omogenitate      | =STDEV (Celinc:Celfin)/AVERAGE (Celinc:Celfin)                      | Calculează omogenitatea unui șir de date                                     |
| Testul Z                       | =ZTEST (Celinc:Celfin, medpop)                                      | Calculează eroarea pentru compararea mediei unui grup cu o valoare dată      |
| Testul t                       | =TTEST (Celinc1:Celfin1,Celinc2:Celfin2, valtail, s, valtype)       | Calculează eroarea pentru a compara mediilor dintre două grupuri de subiecți |
| Testul $\chi^2$ (hi-pătrat)    | =CHITEST (Celinc1:Celfin1, Celinc2:Celfin2)                         | Calculează eroarea pentru a compara diferențele între două distribuții       |

| <b>Denumire funcție</b>   | <b>Forma funcției</b>                      | <b>Unde se poate folosi funcția</b>                                                                               |
|---------------------------|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| Coeficientul de corelație | =CORREL (Celinc1:Celfin1, Celinc2:Celfin2) | Calculează coeficientul de corelație între două variabile                                                         |
| Valoarea tabelară a lui t | =TINV (ns,gl)                              | Calculează valoarea tabelară a lui t în funcție de nivelul de semnificabilitate (ns) și gradele de libertate (gl) |

## 5.2. Utilizarea SPSS

Atunci când se vorbește despre cercetarea în sociologie și psihologie, mai întotdeauna se amintește de prelucrarea datelor cu ajutorul SPSS-ului. Ce este SPSS-ul?

Denumirea provine din engleză (*Statistical Package for the Social Sciences*): „pachet statistic pentru științele sociale”.

Înainte de a începe propriu-zis prezentarea acestui pachet de programe, aș dori să atrag atenția asupra anumitor aspecte

– în acest capitol voi realiza o introducere la un nivel de bază a SPSS-ului.

– pentru o înțelegere mai bună a SPSS-ului se recomandă reluarea unor noțiuni de statistică.

– trebuie să înțelegeți că acest capitol acoperă doar o parte din lucrurile ce se pot realiza cu SPSS-ul. De aceea, după ce veți înțelege bine noțiunile predate, recomandăm să încercați și alte metode statistice ale acestui program.

### 5.2.1. Ferestre și fișiere în SPSS

În SPSS veți avea ocazia să lucrați cu trei tipuri de ferestre, având posibilitatea să salvați conținutul fiecărei ferestre

1 – prima fereastră, pe care o veți întâlni, va fi fereastra **Data Editor**<sup>99</sup>. Are rolul de a defini și introduce datele și, în același timp, de a ajuta la realizarea unor proceduri statistice.

2 – fereastra **Output**<sup>100</sup> – are rolul de a afișa rezultatele testelor statistice.

3 – fereastra **Syntax** – are rolul de a memora operațiile realizate asupra datelor. Aceasta fereastră se va deschide automat atunci când veți da comanda **Paste**.

Oricare din aceste ferestre poate fi salvată. Pentru a executa această comandă trebuie, mai întâi, activată fereastra ce dorim s-o salvăm, apoi se execută o succesiune de comenzi asemănătoare cu salvarea unui document Word.

Așa cum la salvarea unui document în Word, acesta va primi automat extensia .doc, la fel și aici fiecare fereastră după scrierea numelui dat va avea și o extensie cu ajutorul căreia putem să știm ce tip de fereastră avem. În tabelul de mai jos prezentăm ferestrele și extensiile corespunzătoare:

| <i>Numele ferestrei</i> | <i>Extensia fișierului salvat</i> |
|-------------------------|-----------------------------------|
| Editor                  | .sav                              |
| Output                  | .spo                              |
| Syntax                  | .sps                              |

### 5.2.2. Introducerea datelor

În lucrul cu SPSS-ul avem două moduri de introducere a datelor: de la tastatură sau dintr-un program cu foi de lucru<sup>101</sup>. În cadrul acestei lucrări ne vom referi la primul mod de lucru. Deschizând sesiunea de lucru cu SPSS pentru Windows, pe ecran va apărea o fereastră în care vom fi întrebați în legătură cu ceea ce dorim să realizăm:

---

<sup>99</sup> Editarea datelor.

<sup>100</sup> De ieșire.

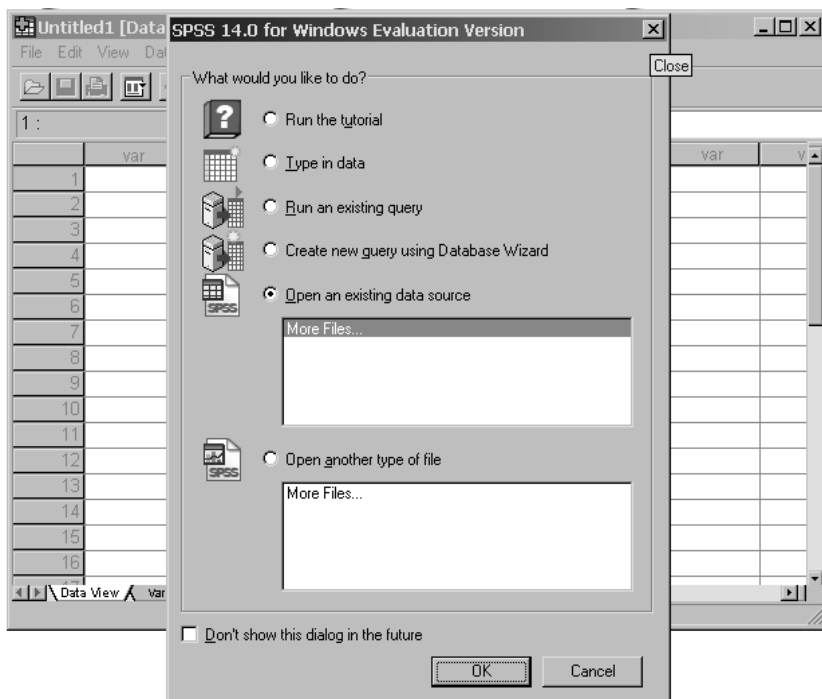
<sup>101</sup> Gen Excel.

---

<sup>99</sup> Editarea a datelor.

<sup>100</sup> De ieșire.

<sup>101</sup> Gen Excel.



Având în vedere denumirea secțiunii, vom alege opțiunea de introducere de date noi<sup>102</sup> după care vom da un clic pe butonul OK. Atunci fereastra de dialog va dispărea și va rămâne o foaie gen Excel în care putem introduce datele. Singura diferență este că în această fereastră coloanele nu sunt notate cu litere, ci la fiecare coloană avem aceeași etichetă *var*. Datele se introduc de la tastatură și după scriere se apasă tasta *Enter*.

Atunci când vom completa acest tabel, la ca și în fel Excel, atunci când introducem datele trebuie să punem pe fiecare linie câte un caz (de exemplu, atunci când dorim să analizăm un lot de 31 de subiecți, fiecare subiect va fi pus pe o linie cu toate caracteristicile lui) și pe fiecare coloană caracteristicile corespunzătoare fiecărui caz (în exemplul de dinainte, dacă dorim

---

<sup>102</sup> *Type in data*

---

<sup>102</sup> *Type in data.*

să realizăm un studiu în funcție de vârstă, mediu, ocupație etc. atunci vom avea o coloană pentru vârstă, una pentru mediu, alta pentru ocupație etc.). De asemenea se recomandă ca fiecărui subiect să-i atribuim un număr<sup>103</sup>. Pentru identificarea sexului subiecților se poate atribui valoarea 1 pentru masculin și 2 pentru feminin.

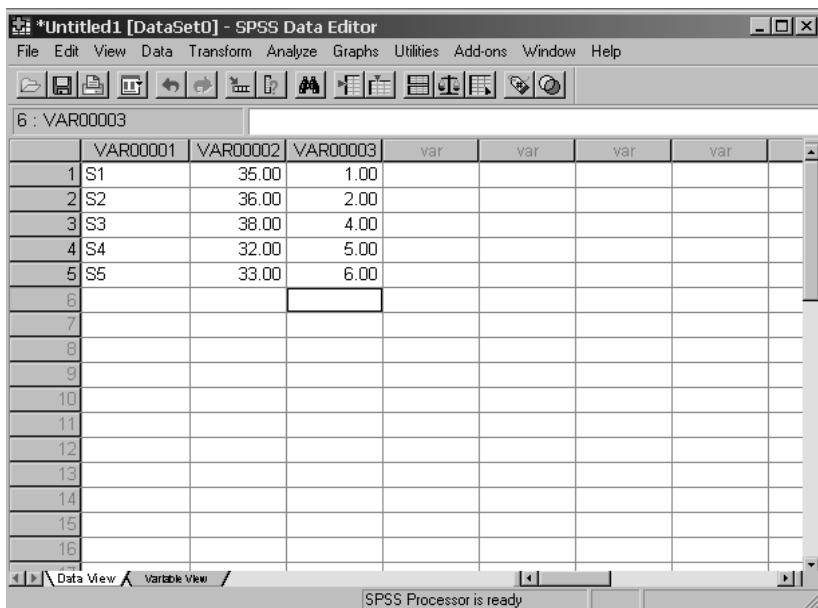
### 5.2.3. Crearea unui fișier de date

În această secțiune vom prezenta modul de realizare a unui fișier de date pas cu pas. Precizăm că operațiile pe care le vom prezenta sunt specifice versiunii SPSS 14.0, în versiuni anterioare procedeul diferind la anumite etape de creare a fișierului de date.

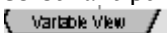
În fereastra în care am introdus datele, în partea de stânga jos avem două butoane: . Implicit este activ cel cu „Data View”, adică fereastra unde se introduc datele. Să presupunem că am introdus următoarele date:

---

<sup>103</sup> Dar nu este obligatoriu.



<sup>103</sup> Dar nu este obligatoriu.

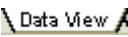
Pentru a defini baza de date va trebui să specificăm pentru fiecare coloană tipul datelor. Pentru aceasta vom da un clic pe butonul  . Va apărea o nouă fereastră:

|    | Name     | Type    | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|----|----------|---------|-------|----------|-------|--------|---------|---------|-------|---------|
| 1  | VAR00001 | String  | 2     | 0        |       | None   | None    | 9       | Left  | Nominal |
| 2  | VAR00002 | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |
| 3  | VAR00003 | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |
| 4  |          |         |       |          |       |        |         |         |       |         |
| 5  |          |         |       |          |       |        |         |         |       |         |
| 6  |          |         |       |          |       |        |         |         |       |         |
| 7  |          |         |       |          |       |        |         |         |       |         |
| 8  |          |         |       |          |       |        |         |         |       |         |
| 9  |          |         |       |          |       |        |         |         |       |         |
| 10 |          |         |       |          |       |        |         |         |       |         |
| 11 |          |         |       |          |       |        |         |         |       |         |
| 12 |          |         |       |          |       |        |         |         |       |         |
| 13 |          |         |       |          |       |        |         |         |       |         |
| 14 |          |         |       |          |       |        |         |         |       |         |
| 15 |          |         |       |          |       |        |         |         |       |         |
| 16 |          |         |       |          |       |        |         |         |       |         |
| 17 |          |         |       |          |       |        |         |         |       |         |

În prima coloană avem eticheta *Name*. Aceasta reprezintă lista etichetelor ce apar în foaia *Data View* pentru fiecare coloană. Prima coloană din baza de date este cea cu etichetele pe care le-am dat fiecărui subiect. Deci, vom denumi aceasta coloană *Etichete*. Pentru aceasta vom da un clic pe VAR00001 și de la tastatură vom scrie „Etichete”, după ce vom da *Enter* în locul lui VAR00001 va apărea denumirea „Etichete”. În următoarea celulă din dreapta trebuie să definim tipul datelor ce se vor introduce pe coloana „Etichete”. Deoarece datele introduse în acea coloană conțin litere, automat este aleasă opțiunea „String”. Deci, putem lăsa opțiunea neschimbată. La opțiunea „Width” se specifică numărul maxim de caractere ce se pot avea datele din aceste coloane. Implicit a fost aleasă valoarea 2, două caractere, deoarece din datele introduse nu avem decât etichete de două caractere. Dacă dorim să schimbăm și să putem introduce date de lungime mai mare, dăm un clic pe valoarea 2 și scriem de la tastatură numărul 8.

Pe liniile următoare vom scrie la etichete test 1, respectiv test 2 și lăsăm neschimbate celelalte opțiuni. Vom obține fereastra:

|   | Name     | Type    | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|---|----------|---------|-------|----------|-------|--------|---------|---------|-------|---------|
| 1 | Etichete | String  | 2     | 0        |       | None   | None    | 9       | Left  | Nominal |
| 2 | Test01   | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |
| 3 | Test02   | Numeric | 8     | 2        |       | None   | None    | 8       | Right | Scale   |

iar dacă dăm un clic pe butonul  vom obține fereastra:

|   | Etichete | Test01 | Test02 | var |
|---|----------|--------|--------|-----|
| 1 | S1       | 35.00  | 1.00   |     |
| 2 | S2       | 36.00  | 2.00   |     |
| 3 | S3       | 38.00  | 4.00   |     |
| 4 | S4       | 32.00  | 5.00   |     |
| 5 | S5       | 33.00  | 6.00   |     |
| 6 |          |        |        |     |
| 7 |          |        |        |     |

Ca un rezumat al celor explicate mai sus vom da un tabel cu proprietățile ce se pot specifica în legătură cu fiecare coloană:

| <i>Tipul de comandă</i>                                                                           | <i>Valoarea implicită</i> |
|---------------------------------------------------------------------------------------------------|---------------------------|
| <b>Type</b> – permite definirea tipului variabilei (numeric, șir de numere, logic)                |                           |
| <b>Width</b> – total numere de caractere                                                          | 8                         |
| <b>Decimals</b> (în cazul datelor numerice) – numărul de zecimale cu care dorim să scriem numărul | 2                         |
| <b>Missing Values</b> – permite desemnarea anumitor scoruri ce pot lipsi la colectarea datelor    |                           |
| <b>Columns</b> – permite schimbarea maximului de caractere dintr-o coloană                        | 8                         |
| <b>Align</b> – permite desemnarea alinierii într-o coloană                                        | dreapta                   |
| <b>Measure</b> – permite desemnarea tipului de date cu care lucrăm <sup>104</sup>                 | scale                     |

---

<sup>104</sup> De exemplu, dacă avem date de tip șir de caractere atunci se va selecta opțiunea *Nominal*.

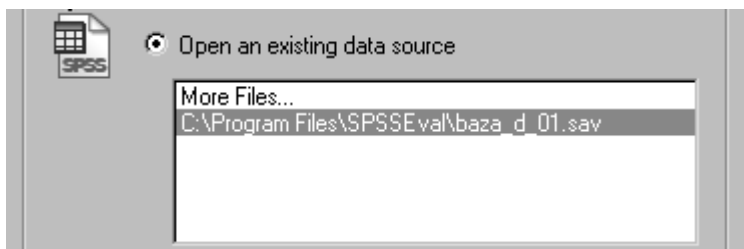
---

<sup>104</sup> De exemplu, dacă avem date de tip șir de caractere atunci se va selecta opțiunea *Nominal*.

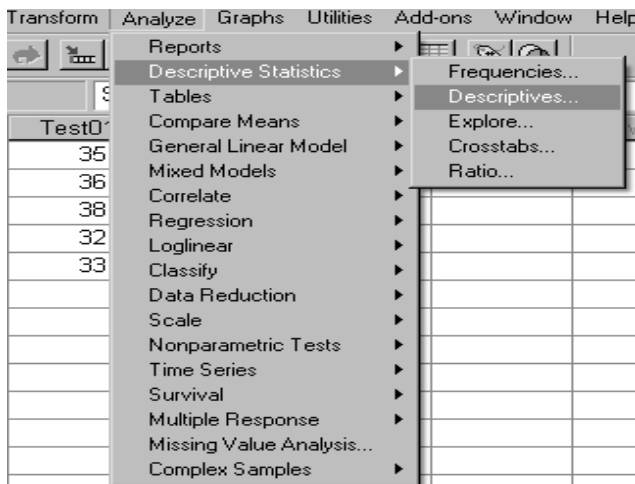
În continuare vom salva această bază de date. Salvarea se realizează asemănător cu modul în care se realizează salvarea în Excel, adică dăm un clic pe opțiunea *File*, apoi din lista ce va apărea vom selecta opțiunea *Save*. Când salvăm prima oară va trebui să indicăm locul unde salvăm baza de date și denumirea bazei de date. Să presupunem că am salvat această bază de date cu denumirea *Baza\_d\_01*. La salvarea bazei de date putem da o opțiune de salvare a bazei de date ca fișier din alt program, de exemplu Excel.

#### 5.2.4. Statistica descriptivă a bazei de date

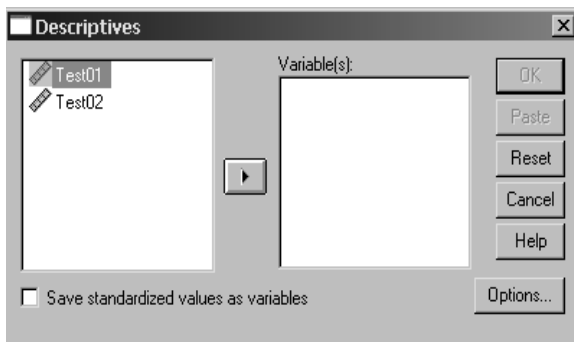
Odată realizată baza de date corect, putem obține o serie de informații utile din aceste date. Un prim lucru pe care-l putem obține este o statistică descriptivă a datelor (număr, medie, abatere standard). Pentru aceasta deschidem baza de date (în fereastra ce apare când deschidem SPSS-ul alegem opțiunea *Open an existing data source*).



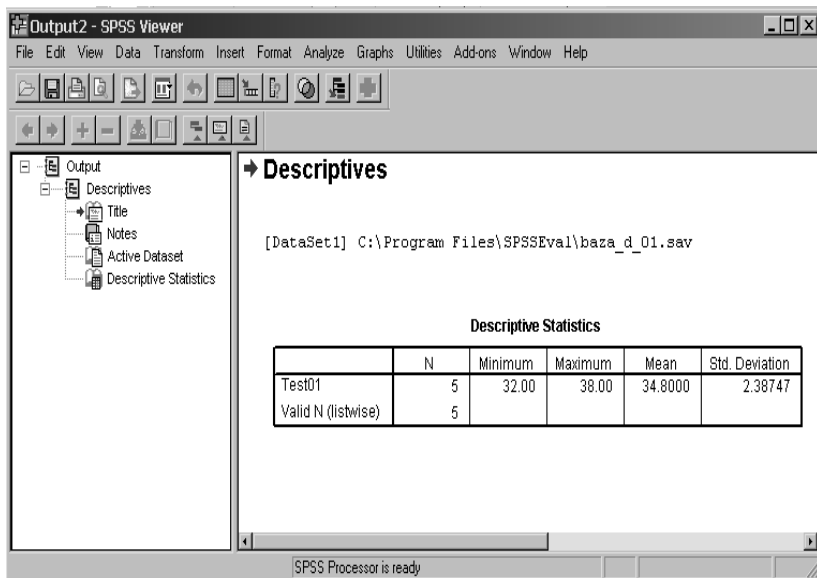
Iar din lista de baze de date alegem *baza\_d\_01.sav*. Se va deschide fereastra cu baza de date și din meniul acestei ferestre vom da un clic pe opțiunea *Analyze*, apoi pe *Descriptive statistics* și apoi vom da un clic pe opțiunea *Descriptives* ca în imaginea de mai jos



Se va deschide fereastra de mai jos:



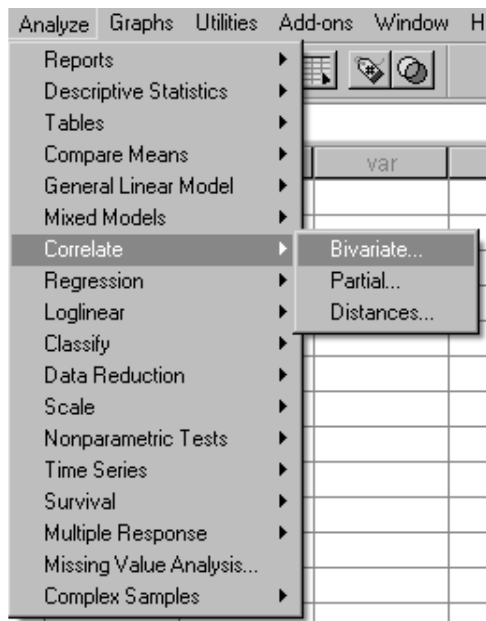
Să presupunem că dorim să obținem statistica descriptivă pentru datele de la testul 01. Dăm un clic pe „Test01”, apoi un clic pe butonul , după care, în fereastra denumită „Variables (s)”, va apărea scris „Tes01” și butonul  se va activa. Dând clic pe acest buton, vom obține fereastra cu datele statisticii descriptive referitoare la datele șirului numit „Test01”



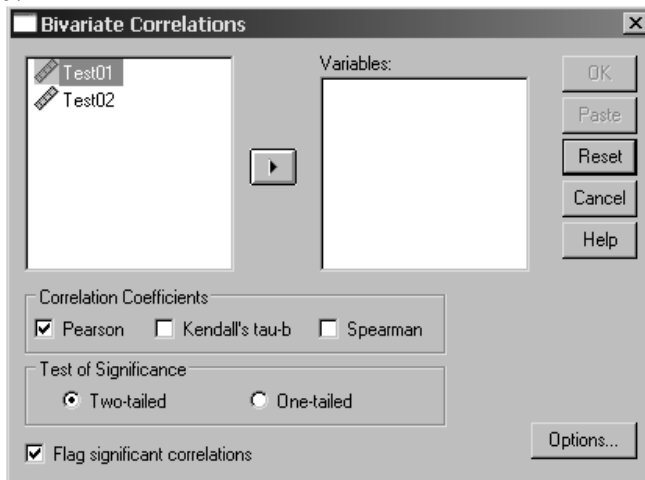
Rezultatele se pot salva sau se pot tipări urmând aceleași operații ca la salvarea unui document în Word.

### 5.2.5. Corelația

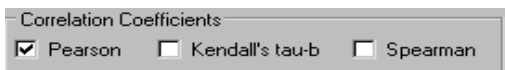
Să deschidem din nou baza de date, așa cum am prezentat la începutul secțiunii 5.2.4. Dorim să calculăm coeficientul de corelație dintre cele două teste din baza de date. Pentru aceasta, din meniu, vom selecta opțiunea *Analyze*, apoi *Correlate* și vom merge pe opțiunea *Bivariate*, ca în imaginea de mai jos:



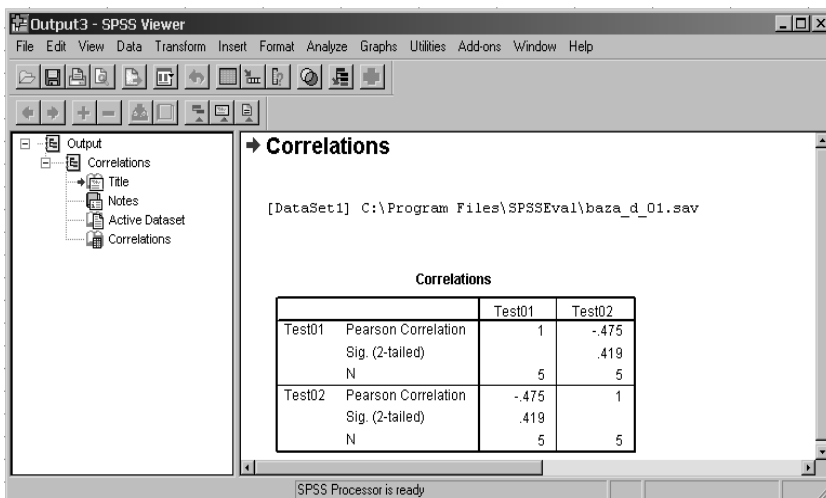
După ce vom da clic pe opțiunea *Bivariate* va apărea fereastra:



Dăm un clic pe *Test01* și apoi pe , procedând analog cu *Test02*. În final, cele două vor apărea pe fereastra din dreapta, butonul OK va deveni activ. Verificăm să avem bifată opțiunea *Pearson* din fereastră



și opțiunea *Two-tailed* care se află pe fereastră mai jos. Apoi dăm un clic pe butonul . Vor apărea rezultatele:



**Correlations**

[DataSet1] C:\Program Files\SPSSEval\baza\_d\_01.sav

|        |                     | Test01 | Test02 |
|--------|---------------------|--------|--------|
| Test01 | Pearson Correlation | 1      | -.475  |
|        | Sig. (2-tailed)     |        | .419   |
|        | N                   | 5      | 5      |
| Test02 | Pearson Correlation | -.475  | 1      |
|        | Sig. (2-tailed)     | .419   |        |
|        | N                   | 5      | 5      |

SPSS Processor is ready

În partea dreaptă a ferestrei vom avea afișate rezultatele. Coeficientul de corelație, *Pearson Correlation*, are valoarea -0,475, iar eroarea  $p = 0,419$ . Ca să avem un coeficient de corelație semnificativ ar trebui ca  $p < 0,05$  și cum  $p$ -ul obținut  $p = 0,419 > 0,05$ , rezultă că între cele două teste nu există o corelație semnificativă.

### 5.2.6. Comparația între mediile a două loturi

În exemplul folosit în secțiunile anterioare nu se poate folosi comparația între medii, deci vom realiza o altă bază de date.

Să presupunem că pe două loturi de subiecți aplicăm un test *Test03*. Pentru claritatea expunerii să presupunem că loturile sunt formate din doar 10 subiecți fiecare.

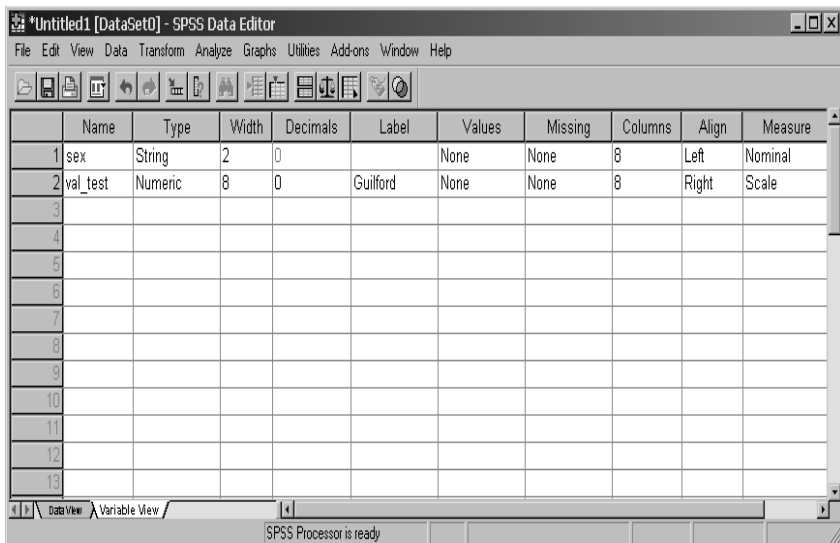
#### **Exemplul 82**

În urma aplicării testului Guilford de abilități ale gândirii divergente (flexibilitatea) la un liceu s-au obținut următoarele rezultate:

| <i>Băieți</i> | <i>Fete</i> |
|---------------|-------------|
| 5             | 6           |
| 11            | 5           |
| 11            | 5           |
| 8             | 6           |
| 7             | 3           |
| 6             | 6           |
| 13            | 8           |
| 6             | 6           |
| 8             | 7           |
| 11            | 8           |

De cele mai multe ori erorile apar de la introducerea datelor. Influențați de modul în care se introduc datele în Excel există tendința de a le pune în același mod în baza de date din SPSS. Fiind vorba de o bază de date, punerea pe aceeași linie a valorilor unui băiat cu valoarea unei fete (de exemplu, pentru primele 2 valori la băieți 5 și la fete 6) ar putea da impresia că între băiatul și fata care sunt pe o linie ar putea să existe o legătură. De fapt, avem 20 de subiecți care nu au legătură unul cu altul, loturile sunt independente, și pentru fiecare subiect în baza de date specificăm sexul și valoarea obținută la test. De aceea, în baza de date din SPSS, pe prima coloana

vom avea declarația sexului subiectului, iar pe a doua coloană vom pune valoarea obținută la testul Guilford. În același timp, vom realiza și o codificare: vom nota cu *m* băieții și cu *f* fetele. Deci, în foaia de declarație a variabilelor vom scrie ca în figura de mai jos:



La variabila *val\_test* am pus în dreptul coloanei *Label* eticheta *Guilford* pentru ca pe tabelul cu rezultate să apară și denumirea testului folosit. Apoi, în foaia *Data View*, vom introduce datele astfel încât pe coloana denumită *Sex* să avem *m* sau *f* corespunzătoare băieților, respectiv fetelor. În urma completării vom obține următorul tabel:

**\*Untitled1 [DataSet0] - SPSS Data Editor**

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

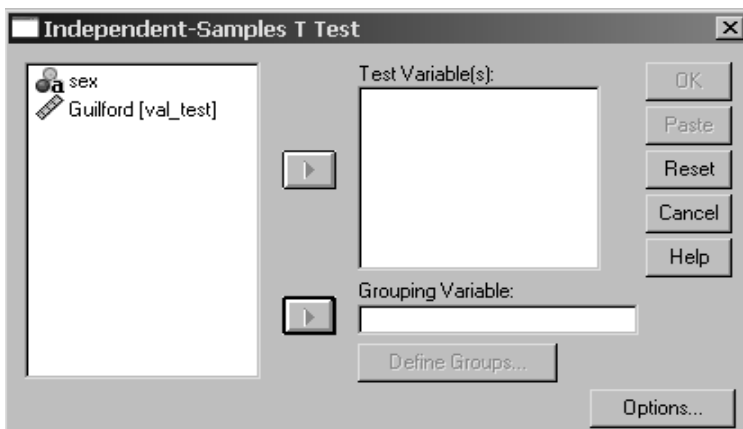
20 :

|    | sex | val_test | var | var | var | var |
|----|-----|----------|-----|-----|-----|-----|
| 1  | m   | 5        |     |     |     |     |
| 2  | m   | 11       |     |     |     |     |
| 3  | m   | 11       |     |     |     |     |
| 4  | m   | 8        |     |     |     |     |
| 5  | m   | 7        |     |     |     |     |
| 6  | m   | 6        |     |     |     |     |
| 7  | m   | 13       |     |     |     |     |
| 8  | m   | 6        |     |     |     |     |
| 9  | m   | 8        |     |     |     |     |
| 10 | m   | 11       |     |     |     |     |
| 11 | f   | 6        |     |     |     |     |
| 12 | f   | 5        |     |     |     |     |
| 13 | f   | 5        |     |     |     |     |
| 14 | f   | 6        |     |     |     |     |
| 15 | f   | 3        |     |     |     |     |
| 16 | f   | 6        |     |     |     |     |
| 17 | f   | 8        |     |     |     |     |
| 18 | f   | 6        |     |     |     |     |
| 19 | f   | 7        |     |     |     |     |
| 20 | f   | 8        |     |     |     |     |

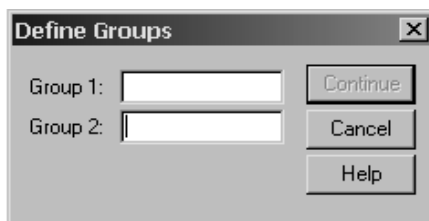
Data View Variable View

SPSS Processor is ready

Deci, până la linia a 10-a inclusiv, am introdus valorile corespunzătoare băieților, iar din linia a 11-a la linia a 20-a am introdus valorile fetelor. Pentru a calcula testul t pentru loturi independente vom da un clic pe opțiunea *Analyze* din meniu, apoi pe *Compare Means*, iar din fereastra ce se va deschide vom da un clic pe opțiunea *Independent-Samples T Test*. După executarea acestei succesiuni de comenzi va apărea o fereastră ca-n imaginea de mai jos:



Variabila *sex* va fi trecută la *Grouping Variable*, iar variabila *Guilford (val\_test)* va fi trecută la *Test Variable (s)*. Mai apoi trebuie să precizăm notațiile pentru grupuri. Pentru aceasta vom da un clic pe butonul **Define Groups...** și se va deschide fereastra:



În dreptunghiul corespunzător lui *Group 1* vom scrie *m*, iar în celălalt dreptunghi vom scrie *f*. Atunci butonul *Continue* va deveni activ. Dăm un clic pe el și în dreptul variabilei *sex* va apărea scris ('*m*' '*f*'). Acum putem da comanda de calculare a testului *t* și dând un clic pe butonul *OK* va apărea următoarea fereastră de rezultate:

## → T-Test

[DataSet0]

Group Statistics

| sex        | N  | Mean | Std. Deviation | Std. Error Mean |
|------------|----|------|----------------|-----------------|
| Guilford m | 10 | 8.60 | 2.716          | .859            |
| f          | 10 | 6.00 | 1.491          | .471            |

Independent Samples Test

|          |                             | Levene's Test for Equality of Variances |      | t-test for Equality of Means |        |                 |                 |                       |                                           |       |
|----------|-----------------------------|-----------------------------------------|------|------------------------------|--------|-----------------|-----------------|-----------------------|-------------------------------------------|-------|
|          |                             | F                                       | Sig. | t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |       |
|          |                             |                                         |      |                              |        |                 |                 |                       | Lower                                     | Upper |
| Guilford | Equal variances assumed     | 6.946                                   | .017 | 2.654                        | 18     | .016            | 2.600           | .980                  | .542                                      | 4.658 |
|          | Equal variances not assumed |                                         |      | 2.654                        | 13.971 | .019            | 2.600           | .980                  | .498                                      | 4.702 |

Se observă că se realizează în partea de sus a ferestrei o statistică descriptivă a datelor în tabelul denumit *Group Statistics*, iar rezultatele testului vor fi puse în tabelul denumit *Independent Samples Variables*. Tabelul are trei mari părți: în prima parte sunt puse etichetele

|          |                             |
|----------|-----------------------------|
|          |                             |
| Guilford | Equal variances assumed     |
|          | Equal variances not assumed |

În a doua parte se calculează testul pentru studiul egalității varianțelor:

| Levene's Test for Equality of Variances |      |
|-----------------------------------------|------|
| F                                       | Sig. |
| 6.946                                   | .017 |

Pe exemplul prezentat am obținut valoarea  $F = 6,946$  și o valoare  $p = 0,017$ . Deoarece  $p < 0,05$  se poate presupune că cele două loturi au varianțe egale. În cea de-a treia parte se dau datele rezultate direct din aplicarea testului  $t$ :

| t-test for Equality of Means |        |                 |                 |                       |                                           |       |
|------------------------------|--------|-----------------|-----------------|-----------------------|-------------------------------------------|-------|
| t                            | df     | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference |       |
|                              |        |                 |                 |                       | Lower                                     | Upper |
| 2.654                        | 18     | .016            | 2.600           | .980                  | .542                                      | 4.658 |
| 2.654                        | 13.971 | .019            | 2.600           | .980                  | .498                                      | 4.702 |

Având în vedere rezultatul testului de comparare a varianțelor, vom lua în considerare doar rezultatele de pe prima linie. În prima coloană avem valoarea  $t = 2,654$ , în a doua coloană avem gradele de libertate  $df = 18$ , în a treia coloană avem valoarea  $p = 0,016$ . Pentru ca  $p < 0,05$  înseamnă că avem diferențe semnificative între medii celor două loturi, deci putem spune că media băieților  $m_b = 8,60$  este semnificativ mai mare ca media fetelor  $m_f = 6,00$ .

### 5.2.7. Testul $\chi^2$ (hi-pătrat)<sup>105</sup>

În secțiunea 5.2. am văzut că modul de a completa o bază de date în SPSS este diferit de modul în care aceste date se pun într-o foaie de lucru, de exemplu în Excel. Ideea de bază este înregistrarea și faptul că pe fiecare linie trebuie pusă câte o înregistrare cu toate componentele ei ce vor fi folosite în analiza datelor. Dacă, în cazul exemplului din secțiunea anterioară, aveam pentru fiecare subiect de pus sexul și valoarea testului, în cazul aplicării testului hi-pătrat în SPSS avem de făcut trei declarații. Pentru claritatea expunerii vom folosi un exemplu:

#### Exemplul 87

Într-un centru de reeducare s-a pus băieților următoarea întrebare: ***Din ce cauză nu-ți place școala?*** În urma centralizării datelor s-au obținut următoarele rezultate:

|                         | 16-17 ani | 18-22 ani |
|-------------------------|-----------|-----------|
| îmi este greu să învăț  | 1         | 4         |
| mă plictisesc           | 18        | 11        |
| mă enervează profesorii | 7         | 1         |
| nu-mi folosește         | 10        | 19        |

Dorim să studiem dacă există diferențe semnificative între cele două loturi de subiecți.

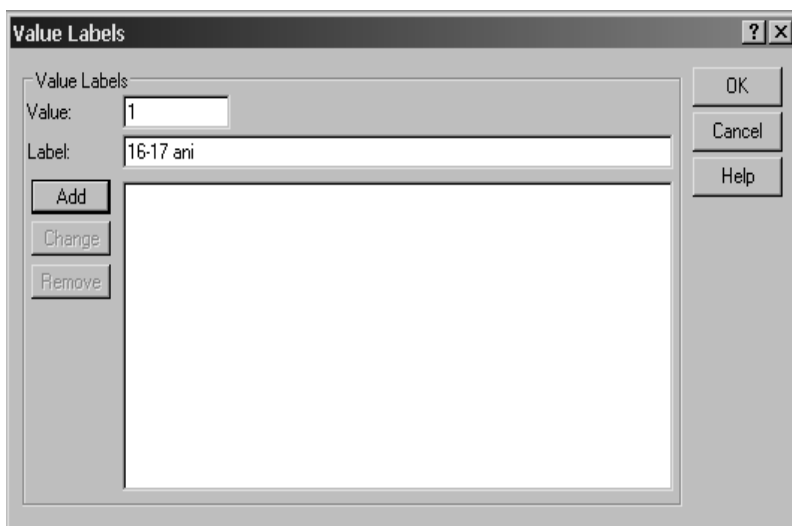
Pentru început vom completa foaia cu *Variable View* a bazei de date din SPSS. Vom avea trei tipuri de caracteristici: *Grup\_vârstă*, *var\_răspuns* și *frecvențe*. Vom face și următoarele codificări:

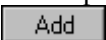
| Grup vârstă        |            | Variante de răspuns     |            |
|--------------------|------------|-------------------------|------------|
| Denumire categorie | Codificare | Denumire categorie      | Codificare |
| 16-17 ani          | 1          | îmi este greu să învăț  | 1          |
| 18-22 de ani       | 2          | mă plictisesc           | 2          |
|                    |            | mă enervează profesorii | 3          |
|                    |            | nu-mi folosește         | 4          |

<sup>105</sup> În engleză – *chi-test*

<sup>105</sup> În engleză – *chi-test*

Deschidem *Data editor* și definim, mai întâi, baza de date în foaia *Variable View*. Pentru ca în tabelul cu rezultate să apară și denumirile, nu doar codurile, în tabelul de declarații, în coloana *Values*, vom scrie și etichetele corespunzătoare codurilor. De exemplu, pentru categoria *Grup vârstă* dăm un clic pe celula din dreptul liniei categoriei și coloana *Values*. Atunci, în dreapta celei, va apărea butonul . Dăm un clic pe acest buton și va apărea fereastra:



În dreptul lui *Value* vom scrie codul, iar în dreptul *Label* vom scrie denumirea. Apoi dăm un clic pe butonul . Iar codul și eticheta vor apărea în dreptunghiul de mai jos. Astfel, introducem toate codificările și etichetele corespunzătoare categoriei

*Grup\_varsta*<sup>106</sup>. Analog se vor introduce și codificările pentru *var\_raspuns*. În final, tabelul *Variable View* va arăta astfel:

<sup>106</sup> În coloana *Name* nu se acceptă spații libere, de aceea am pus \_ între cele două cuvinte ce compun denumirea categoriei.

|    | Name        | Type    | Width | Decimals | Label | Values           | Missing | Colu |
|----|-------------|---------|-------|----------|-------|------------------|---------|------|
| 1  | Grup_varsta | Numeric | 8     | 0        |       | {1, 16-17 ani}.. | None    | 8    |
| 2  | Var_raspun  | Numeric | 8     | 0        |       | {1, imi este gre | None    | 8    |
| 3  | Frecvente   | Numeric | 8     | 0        |       | None             | None    | 8    |
| 4  |             |         |       |          |       |                  |         |      |
| 5  |             |         |       |          |       |                  |         |      |
| 6  |             |         |       |          |       |                  |         |      |
| 7  |             |         |       |          |       |                  |         |      |
| 8  |             |         |       |          |       |                  |         |      |
| 9  |             |         |       |          |       |                  |         |      |
| 10 |             |         |       |          |       |                  |         |      |
| 11 |             |         |       |          |       |                  |         |      |
| 12 |             |         |       |          |       |                  |         |      |

Dăm un clic pe *Data View* pentru a introduce datele. Datele se vor introduce asemănător cu modul în care le-am introdus la testul t:

<sup>106</sup> În coloana *Name* nu se acceptă spații libere, de aceea am pus \_ între cele două cuvinte ce compun denumirea categoriei.

**\*Untitled1 [DataSet0] - SPSS Data Editor**

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

19 :

|    | Grup_varsta | Var_raspuns | Frecvente | var | var |
|----|-------------|-------------|-----------|-----|-----|
| 1  | 1           | 1           | 1         |     |     |
| 2  | 1           | 2           | 18        |     |     |
| 3  | 1           | 3           | 7         |     |     |
| 4  | 1           | 4           | 10        |     |     |
| 5  | 2           | 1           | 4         |     |     |
| 6  | 2           | 2           | 11        |     |     |
| 7  | 2           | 3           | 1         |     |     |
| 8  | 2           | 4           | 19        |     |     |
| 9  |             |             |           |     |     |
| 10 |             |             |           |     |     |

Data View Variable View SPSS Processor is ready

Următoarea procedură ce trebuie realizată este ponderarea celulelor. Pentru a realiza aceasta, dăm un clic pe opțiunea *Data* din meniu, iar din fereastra ce se va deschide alegem opțiunea *Weight Cases*<sup>107</sup>. Va apărea fereastra:

**Weight Cases**

☒ Do not weight cases  
☐ Weight cases by

Frequency Variable:

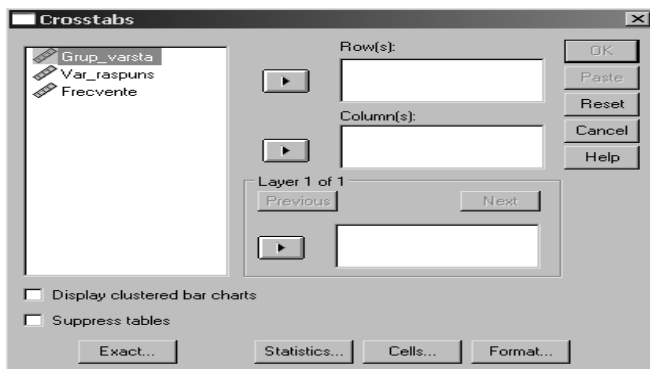
Current Status: Do not weight cases

OK Paste Reset Cancel Help

În această fereastră dăm un clic pe opțiunea *Weight cases by*, se va activa butonul  de la *Frequency Variable*. Dăm un clic pe *Frecvențe* și apoi pe butonul  pentru a specifica coloana unde sunt frecvențele. Apoi dăm un clic pe butonul OK. Odată specificat faptul că lucrăm cu frecvențe, putem trece la calcularea efectivă a testului  $\chi^2$ . Dăm un clic pe opțiunea *Analyse* din meniu și din fereastra

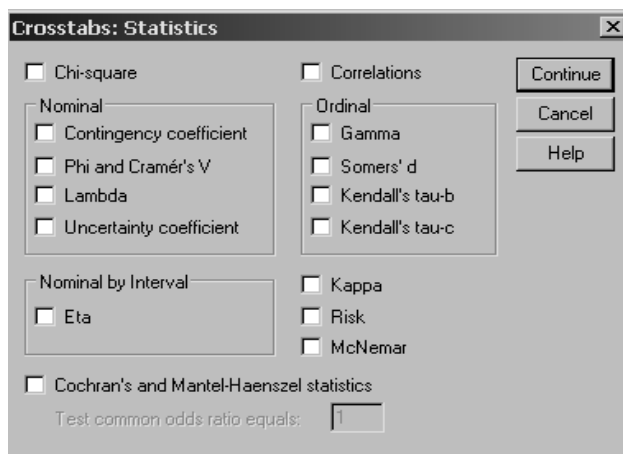
<sup>107</sup> Ultima în listă.

ce se va deschide alegem opțiunea *Descriptive Statistics*. Din noua fereastră ce se va deschide vom alege opțiunea *Crosstabs*. Va apărea următoarea fereastră:

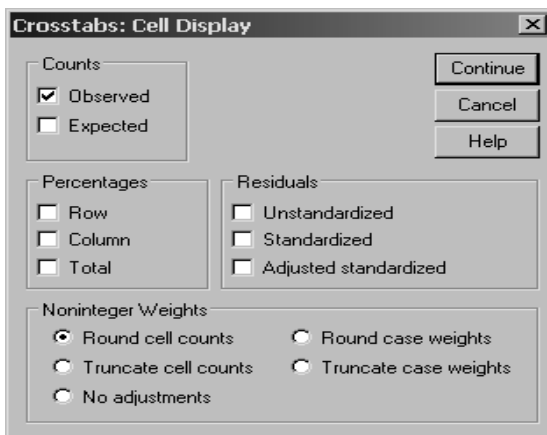


<sup>107</sup> Ultima în listă.

Variabila *Grup\_varsta* o vom pune la *Row (s)*, dând un clic pe butonul , iar variabila *Var\_raspuns* o vom pune la *Column (s)*. Apoi dăm un clic pe butonul **Statistics...** și va apărea fereastra:



Vom da un clic pe *Chi-square* pentru a specifica faptul că vom utiliza testul  $\chi^2$ , apoi dăm un clic pe butonul **Continue** pentru a reveni la fereastra inițială. În fereastra inițială vom da un clic pe butonul **Cells...** iar în fereastra ce se va deschide:



dăm un clic pe opțiunea *Expected* și pe *Unstandardized*, apoi pe butonul **Continue**. Reveniți la fereastra inițială, dăm un clic pe butonul **OK** pentru calcularea rezultatelor testului  $\chi^2$ . În fereastra cu rezultate vor apărea trei tabele.

Primul tabel

**Case Processing Summary**

|                             | Cases |         |         |         |       |         |
|-----------------------------|-------|---------|---------|---------|-------|---------|
|                             | Valid |         | Missing |         | Total |         |
|                             | N     | Percent | N       | Percent | N     | Percent |
| Grup_varsta*<br>Var_raspuns | 71    | 100.0%  | 0       | .0%     | 71    | 100.0%  |

ne indică niște parametrii generali ai datelor introduse în tabel.

Al doilea tabel:

Grup\_varsta \* Var\_raspuns Crosstabulation

|                 |                |                | Var_raspuns               |               |                            |                    | Total |
|-----------------|----------------|----------------|---------------------------|---------------|----------------------------|--------------------|-------|
|                 |                |                | imi este greu<br>sa invat | ma plictisesc | ma enerveaza<br>profesorii | nu-mi<br>foloseste |       |
| Grup_<br>varsta | 16-17 ani      | Count          | 1                         | 18            | 7                          | 10                 | 36    |
|                 |                | Expected Count | 2.5                       | 14.7          | 4.1                        | 14.7               | 36.0  |
|                 |                | Residual       | -1.5                      | 3.3           | 2.9                        | -4.7               |       |
|                 | 18-22 ani      | Count          | 4                         | 11            | 1                          | 19                 | 35    |
|                 |                | Expected Count | 2.5                       | 14.3          | 3.9                        | 14.3               | 35.0  |
|                 |                | Residual       | 1.5                       | -3.3          | -2.9                       | 4.7                |       |
| Total           | Count          |                | 5                         | 29            | 8                          | 29                 | 71    |
|                 | Expected Count |                | 5.0                       | 29.0          | 8.0                        | 29.0               | 71.0  |

ne redă tabelul cu valori. În dreptul liniei, denumită *Residual*, se calculează diferența dintre cele două frecvențe ale celor două grupuri corespunzătoare unei variante de răspuns. De exemplu, la opțiunea *imi este greu sa invat* la grupul de *16-17 ani* la *Residual* se observă valoarea *-1.5*. Aceasta ne arată faptul că pentru această variantă de răspuns cei din categoria *16-17 ani* au mai puține opțiuni decât cei de la categoria *18-22 de ani*.

Al treilea tabel:

Chi-Square Tests

|                                 | Value               | df | Asymp. Sig.<br>(2-sided) |
|---------------------------------|---------------------|----|--------------------------|
| Pearson Chi-Square              | 10.771 <sup>a</sup> | 3  | .013                     |
| Likelihood Ratio                | 11.522              | 3  | .009                     |
| Linear-by-Linear<br>Association | 1.251               | 1  | .263                     |
| N of Valid Cases                | 71                  |    |                          |

a. 4 cells (50.0%) have expected count less than 5. The minimum expected count is 2.46.

ne dă ceea ce rezultă în urma aplicării testului  $\chi^2$ . Valoarea  $\chi^2_{\text{calc}}$  este<sup>108</sup> 10,771 gradele de libertate<sup>109</sup> sunt 3. Valoarea  $p = 0,013$  și cum ea este mai mică de 0,05 înseamnă că avem diferențe semnificative între cele

<sup>108</sup> Pearson Chi-Square

<sup>109</sup> Df

două grupuri, deci putem spune că cei de 16-17 ani au dat răspunsuri semnificativ diferite de cei de 18-22 de ani.

Se poate observa faptul că sub acest tabel ni se atrage atenția că avem 4 celule (50% din date) cu o frecvență foarte mică. Și în continuare se specifică faptul că minimum pentru frecvență este 2,46 (deci frecvența minimă recomandată este 3).

#### 5.2.8. Coeficientul alpha

Atunci când se aplică un instrument ce conține mai mulți itemi, una din problemele ce apar este de a studia dacă itemii tind să măsoare același lucru. Există mai multe metode de a rezolva această problemă, dintre acestea amintim: folosirea coeficientului alpha, metoda split-half etc. În această secțiune vom prezenta un mod de calculare a coeficientului alpha. Cu ajutorul coeficientului alpha se poate rezolva o paletă mai largă de probleme. Dintre acestea amintim:

- Scurtarea unui chestionar prin menținerea sau îmbunătățirea validității interne
- Măsurarea concordanței dintre doi evaluatori diferiți
- Studiul consistenței unei măsuri luate atât la testare, cât și la retestare

În cele ce urmează vom da un exemplu de calcul al unui coeficient specific SPSS-ului.

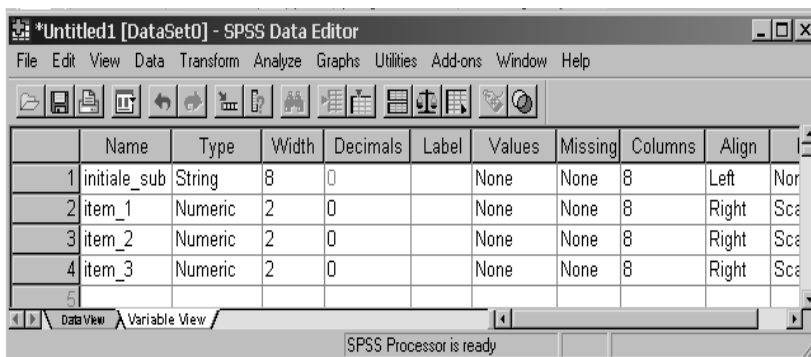
<sup>108</sup> *Pearson Chi-Square*

<sup>109</sup> Df.

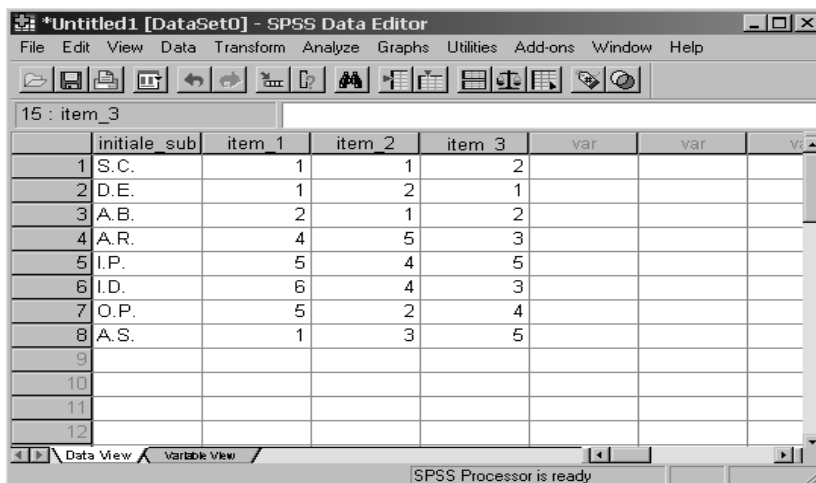
Să presupunem că avem 8 persoane<sup>110</sup> ce răspund la trei itemi. În primul rând, trebuie să definim câmpurile bazei de date. În prima coloană vom trece inițialele subiecților, deci vom defini această coloană ca fiind de tip *String*, iar celelalte coloane ce vor conține valorile obținute pentru fiecare item vor fi de tip *Numeric* de mărime 2 și fără zecimale. După ce am completat aceste date vom obține o foaie ca-n figura de mai jos:

---

<sup>110</sup> Desigur în practică numărul poate fi mult mai mare, dar am optat doar pentru 8 subiecți pentru claritatea expunerii.

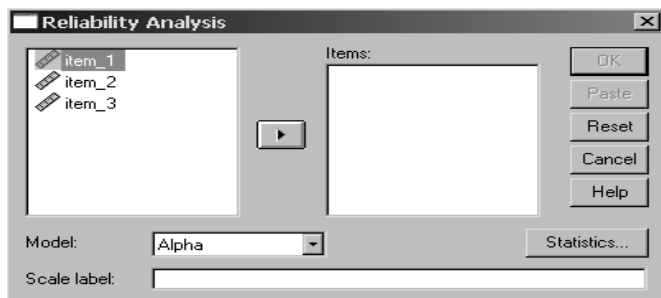


Dăm apoi un clic pe *Data View* și introducem datele ce au reieșit din aplicarea testelor. Va rezulta o foaie ca-n figura de mai jos:

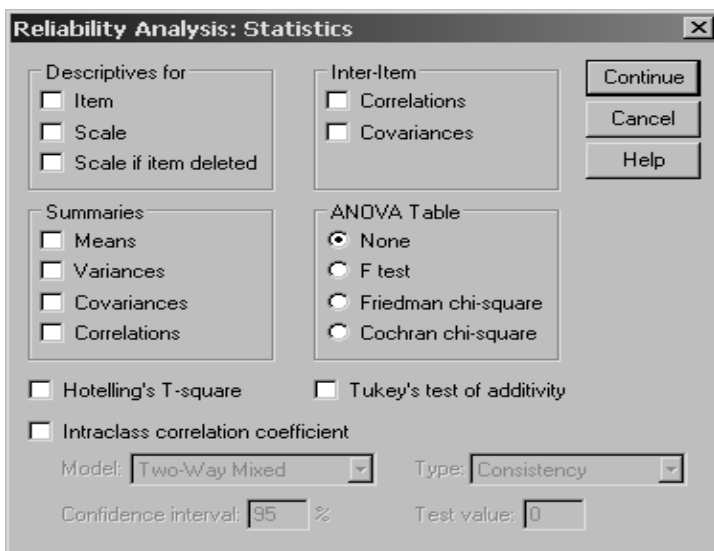


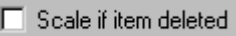
<sup>110</sup> Designul în practică numărul poate fi mult mai mare, dar am optat doar pentru 8 subiecți pentru claritatea expunerii.

Pentru a stabili fidelitatea itemilor, prin metoda coeficientului alpha, selectăm din meniu opțiunea *Analyze*, iar în fereastra ce se va deschide dăm un clic pe opțiunea *Scale* și apoi pe *Reliability Analysis*. Se va deschide fereastra:



Trecem cei trei itemi în fereastra din dreapta cu ajutorul butonului , după care dăm un clic pe butonul . Și în fereastra



Bifăm opțiunea , apoi dăm un clic pe butonul . Se va reveni la fereastra inițială unde se va da un clic pe butonul . În fereastra de rezultate vor apărea trei tabele. Primul tabel prezintă validitatea datelor.

Al doilea tabel

### Reliability Statistics

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .734             | 3          |

Ne dă valoarea coeficientului alpha<sup>111</sup>, care, în acest caz, este 0,734. Se consideră că fidelitatea itemilor este asigurată de la o valoare a coeficientului mai mare de 0,8, deci putem spune că nu avem asigurată validitatea itemilor.

Ultimul tabel

### Item-Total Statistics

|        | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|--------|----------------------------|--------------------------------|----------------------------------|----------------------------------|
| item_1 | 5.88                       | 6.411                          | .594                             | .646                             |
| item_2 | 6.25                       | 9.071                          | .653                             | .559                             |
| item_3 | 5.88                       | 10.411                         | .490                             | .727                             |

ne arată variația coeficientului alpha dacă eliminăm un item. De exemplu, dacă eliminăm itemul 1 obținem un coeficient alpha de 0,646.

### 5.3. Concluzii

În această secțiune am încercat să prezentăm doar o parte din posibilitățile celor două soft-uri de prelucrare statistică a datelor. Pentru cei care doresc să vadă mai multe din posibilitățile acestor soft-uri recomandăm, de exemplu, pentru Excel [Bott], [Pom], iar pentru SPSS recomandăm [How].

---

<sup>111</sup> Se mai numește și *coeficientul lui Cronbach*.

### BIBLIOGRAFIE

---

<sup>111</sup> Se mai numește și *coeficientul lui Cronbach*.

- [Abr] Abric J.C. (2002), *Psihologia comunicării. Teorii și metode*, Editura Polirom, Iași
- [Alex] Alexandrescu P. (2004), *Introducere în statistica socială*, Editura Paralela 45, București
- [Bar] Bârză S. (2004), *Inițiere în internet și realizarea paginilor Web*, Editura Fundației România de Mâine, București
- [Bea] Beaty J. (2004), *Lumea în viziunea lui Peter Drucker*, Editura Teora, București
- [Bott] Bott E. (2001), *Utilizare Microsoft Office 2000*, Editura Teora, București
- [Che1] Chelcea S, Mărginean I., Cauc I (1998), *Cercetarea sociologică*, Editura Destin, Deva
- [Che2] Chelcea S (1975), *Chestionarul în investigația sociologică*, Editura Științifică și Enciclopedică, București
- [Col] Colman A., Pulford B. (2006), *A Crash Course in SPSS for Windows*, Blackwell Publishing, Oxford
- [Die] Diestel R. (2005), *Graph Theory*, Springer-Verlang, New York (electronic edition)
- [Durk] Durkheim E. (2002), *Regulile metodei sociologice*, Editura AntetXX Press
- [Fish] Fisher R.A. (1925), *Statistical methods for research workers*, Oliver and Boyd, London
- [Fota1] Fotache M. (2001), *SQL. Dialecte, DB2, Oracle și Visual FoxPro*, Editura Polirom, Iași
- [Fota2] Fotache M. (2005), *Proiectarea bazelor de date*, Editura Polirom, Iași
- [Gat] Gataulin A., Lica D., Pomohaci C.M. (2002), *Biostatistica intuitivă*, Editura CERES, București

- [How] Howitt D., Cramer D. (2006), *Introducere în SPSS pentru psihologie*, Editura Polirom, Iași
- [Isa] Isaic-Maniu A., Mitruț C., Voineagu V. (2003), *Statistică*, Editura Universitară, București
- [Kin] Kinkoph S. (2004), *Power Point 2003 în imagini*, Editura Teora, București
- [Lic] Lica D., Pomohaci C.M. (1998), *Biostatistica*, U.S.A.M.V.B., București
- [Lin] J.K. Lindsey (2004), *Introduction to applied statistics. A modeling approach*, Oxford University Press Inc., New York
- [Mar] Mărginean I. (1982), *Măsurarea în sociologie*, Editura Științifică și Enciclopedică, București
- [Mat] Matheson D.W., Bruce R.L. Beauchamp K.L. (1970), *Introduction to experimental psychology*, Holt, Rinehart and Winston Inc., New York
- [McC] McCollum J.K., Bănaciu C.S. (2005), *Management de proiect – o abordare practică*, Editura Universitară, București
- [More] Moreno J.I. (1969), *Les fondements de la sociometrie*, PUF, Paris
- [Nast] Năstase P., Mihai F. (2001), *Baze de date – Microsoft Access 2000*, Editura Teora, București
- [Pear] Pearson K., Lee A (1902-1903), *Inheritance of physical characters*, Biometrika 2:357
- [Pom] Pomohaci C.M. (2005), *Noțiuni introductive de utilizare a computerului: Word, Excel, Documente HTML*, Editura Didactică și Pedagogică, București
- [Pom] Pomohaci C.M. (2007), *Informatică utilizată pentru aplicații în sociologie și psihologie*, Editura Fundației România de Măine, București
- [Rad] Radu I și colab. (1993), *Metodologie psihologică și analiza datelor*, Editura Sincron, Cluj
- [Rat] Rateau P. (2004), *Metodele și statisticile experimentale în științele umane*, Editura Polirom, Iași

- [Rot] Rotariu T, Bădescu G., Culic I, Mezei E., Mureșan C., *Metode statistice aplicate în științele sociale*, Editura Polirom, Iași
- [Rua] Ruane J.M. (2004), *Essentials of Research Methods*, Blackwellpublishing, Oxford
- [Rum] Rummel R.J. (1983), *Libertarianism and International Violence*, The Journal of Conflict Resolution 27 (March 1983): 27-71
- [Sne] Snedecor G.W. (1968), *Metode statistice aplicate în cercetările de agricultură și biologie*, Editura Didactică și Pedagogică, București
- [Sta] Stahl H.H. (1974), *Teoria și practica investigațiilor sociale*, Editura Științifică și Enciclopedică, București
- [Stan] Stănășilă O. (1985), *Noțiuni și tehnici de matematică discretă*, Editura Științifică și Enciclopedică, București
- [Ste] R.G.D. Steel, J.H.Torrie (1980), *Principles and procedures of statistics. A biometrical approach*, McGraw-Hill International Editions
- [Sto] Stoica V. (2000), *Metodologia analizei politice*, Editura Fides, Iași
- [Tag] Tagiuri R (1952), *Relational analysis: an extension of sociometry method with emphasis upon social perception*, Sociometry, nr.15
- [Țar] Țarcă M. (1998), *Tratat de statistică aplicată*, Editura Didactică și Pedagogică, București
- [Țiț] Țițan E., Ghiță S., Trandaș C. (2001), *Bazele statisticii*, Editura Meteor Press, București
- [Vel] Velicanu M, Bodea C., Lungu I., Ioniță C., Bădescu G. (1999), *Sisteme de Gestione a Bazelor de Date*, Editura Petrion, București

Redactor: Cosmin COMARNESCU  
Tehnoredactor: Marcela OLARU  
Coperta: Marilena BĂLAN

---

Bun de tipar: 14.02.2008; Coli tipar: 11,25  
Format: 16/61×86

---

Editura Fundației *România de Mâine*  
Bulevardul Timișoara nr.58, București, Sector 6  
Tel./Fax: 021/444.20.91; [www.spiruharet.ro](http://www.spiruharet.ro)  
e-mail: [contact@edituraromaniademaine.ro](mailto:contact@edituraromaniademaine.ro)